

Richard Sprague  
PO Box 1523  
Mercer Island  
WA  
98040  
Sprague@ensemblbio.com

16,200 words

Hackers Guide to the Microbiome

by Richard Sprague

## ***Contents***

### Hackers Guide to the Microbiome\_??

Title Page_	1
Contents_	2
Foreword_	4
Chapter One - The Science_	6
Intro_	7
Microbes everywhere!_	11
In the air_	12
Underground_	14
Microbes_	16
What is a microbe?_	17
Microbes in the gut_	19
Microbes of the skin_	20
A crash course in microbiology_	21
A crash course in botany_	26
What are they doing?_	27
Studying the Human Microbiome_	29
Chapter Two - The Technology_	30
Studying the Microbiome_	31
The old way: culturing_	32
16s rRNA_	33
Gene sequencing_	35
Bioinformatics_	36
Chapter Three - How to Analyze Your Microbiome_	37
How to get your data_	38
The basics: web tools_	42
Analyze with Excel_	44
Comparing samples_	49
Comparison over time_	50
Measuring Diversity_	52
Chapter Four - Experiment Examples_	54
A/B Testing_	55
Hacking my sleep_	56
Visit the jungle_	58
Looking into your mouth_	61
Comparing with food_	63
Chapter Five - Organisms to Watch_	66
Good and Bad_	67
Gut Guardians_	68
Chapter Six - Next Steps_	70
Fermented food_	71
Overselling the microbiome_	75
Annotated References_	76



## ***Foreword***

This book is for curious people with little or no knowledge of biology, but who have a reasonable level of comfort with computer software, either as programmer, or as an end-user comfortable installing and running common applications like Excel. Using plenty of hands-on examples, you'll learn the basics of the human microbiome and how to study and manipulate the microbes inside you.

There are four main sections:

1. **The Science:** an introduction to the field of the microbiome. What makes it so interesting? How do scientists go about studying it? How much is known today and what is likely to be discovered in the future?

2. **Technology.** What instruments and tools are available to study the human microbiome? Microscopes have been around for centuries, but in recent decades a whole laboratory of new tools make it possible to study in much more detail. We'll summarize the major technologies, from gene sequencing to bioinformatics, that let scientists identify and study microbes.

3. **How to analyze your own microbiome:** we'll walk step-by-step through the process of getting and then analyzing data about your own microbiome. Starting with easy-to-understand web tools, we'll walk you through the basics, then show you how to do simple analysis in Excel. For more complicated analysis, we'll introduce simple (free) programming tools in Python and show step-by-step how to use them to study your own microbiome as well as compare it to others.

4. **Organisms to watch:** Once you've begun the process of studying your own microbes, you'll want to learn about some of the common ones and what they do. We'll summarize what scientists already know about these microbes and show you how to design experiments that let you watch for changes over time.

5. **Next steps.** The information in this book will introduce you to the basics, but once you see how fascinating the microbiome can be, no doubt you'll want to go further. We'll show you how to begin growing your own favorite microbes, and then give you a taste of some of the tools that professional biologists use to analyze in much more detail. Finally, we've included an annotated selected reference of the books, web sites, courseware, and other places you can go to learn more.

By the end of the book, you will be well on your way to a hands-on understanding of one of the most exciting areas at the frontier of biology.



## **CHAPTER ONE**

### ***The Science***

## ***Intro***

Humans, like computers, can be broadly described in terms of software and hardware. Our software includes things like language, culture, education, behaviors — the aspects of ourselves that seem most flexible, even arbitrary, but which define us as individuals. There are many ways to improve your software: through education, reading, or interesting conversations with others.

To improve your hardware, you can of course measure various things about yourself: temperature, blood tests, activity levels. Wearable sensors like the Apple Watch are making this much easier, and in the near future you'll see an explosion of new and better ways to measure your body's hardware and with better, more actionable measurement come many ways to modify and optimize your health along the way. All of these new wearable computing devices recognize that you are a unique individual and that we always should treat the patient, not the statistics. So what is the part of you that is fundamentally you?

The obvious place to start is with DNA, and our human genes. I was excited when 23andme (and other services) came out about ten years ago, letting each of us inexpensively see our internal hardware, perhaps to learn more about the things that make us unique, and maybe peer into our futures, to the degree that biological traits like susceptibility to various diseases or conditions is hereditary and unchangeable.

But after a while I realized how little of what we learn about DNA is actionable. If you find you have the "gene for" Alzheimer's disease or Parkinson's, or any of the other traits that have been linked to specific genes, what can you do about it? In nearly every case, the advice is exactly what you should be doing even if you don't have that trait: eat well and get lots of exercise.

### **Fortunately, it's not just our DNA**

In fact, if you could point a scanner at me and pull out everything, all the "hardware" inside me right now, you'd find a curious fact: although sure, 99% of the weight and size of what you see is human (blood, skin, bones, organs, etc.), only about 10% of the cells are human, and even less than that — only about 1% — of the DNA-storing genes are human.

What's the rest? Who am I, if only 1% of the genes inside me are human? The answer is microbes, and as befits something that so outnumbers the "human" part of us, they play a large role in everything about what we do, from our health to our moods and even, as I'll show, to our motivations.

### **Microbes are older than any of us**

We tend to think of the invisible, single-celled microbes as "primitive", not nearly as "advanced" as we humans, with our marvelous brains and ability to transform the earth with airplanes and skyscrapers and nuclear reactors and all

the rest. But that's what we would think, wouldn't we? In fact, the microbes are everywhere, literally everywhere on earth, in the sky, even deep underground. We can't go anywhere without encountering microbes because, well, there are even microbes on and inside us. Humans can't survive without microbes. So what does it even mean to say we're "better" or more "advanced" than they are?

Before the microscope, people didn't even know that microbes existed. Similarly, until the advent of large-scale gene sequencing machines in the past ten years, almost nothing was known about the amazing ubiquity and resilience of microbes.

Yes, they're single-celled and yes many of their cellular functions seem more straightforward than the functions of a multi-cellular creature like us, but it would be a mistake to assume that means microbes -- collectively -- are less influential, and certainly it would be a big error to assume they are less important. Fact is, these organisms have been around, breathing, eating, multiplying, for billions of years, often in pretty much the same form that they are today. These things have survived every imaginable planetary condition from volcanoes to the depths of the ocean to the inside of nuclear reactors. Global Warming means nothing to these guys, who have seen and thrived all over the earth since the day life first appeared.

### **What they want**

Because they have such a huge advantage over us, in lifespan (each microbe is an exact divided copy of itself, going back a zillion generations), in speed of replication (they can double in just a few minutes under the right conditions), and ubiquity (as I said, cellwise they far outnumber us), they can afford to colonize every new imaginable environment.

And that's what they do, every time a new frontier opens to them. The moment of your birth, for example, when you left the (mostly) sterile conditions of your mother's womb, they immediately flooded all over your skin, mostly coming from your mother, and in that fresh environment they used their first-mover advantage to get a stronghold that in many cases lasts your entire life. Many (most?) of the microbes that matter arrived inside you this way, originally, and many of them are still there today, decades, even half a century or more later.

To survive, they need one thing: something to eat. Being so tiny, they don't need much, and they mostly eat things that you (and other larger creatures) weren't interested in anyway. (Or, since they were here first, it's probably more accurate to say that you and I must live on the foods that they don't want. A cheeseburger is only food for you because you snatch it faster than they do. Leave it outside for a while and they'll get it eventually).

### **Who's in charge?**



Collectively, the multitude species of microbes will eat just about everything, but individually each species has its preferences. When they're outside the body, as I said, they can "colonize" new territories (like fresh baby skin) to get what they want, but those inside your tummy are at the mercy of whatever it is you decide to put into your mouth.

Often, that's not a big deal: many species thrive on the same proteins, carbohydrates, and fats that you do. But some species do better than others with certain types of foods, and this is where the line between your human desires and theirs becomes unclear.

Eighty percent of all your brain's outside receptors -- counting all the nerve endings everywhere on your skin -- eighty percent complete their connections in the gut. The main switching grounds, an area called the vagus nerve, does something. What? We know very little, but we see some evidence that the purpose -- the reason that not 1% or 10% or 50% but a full 80% of all the receptors go to the gut -- is so the microbes can tell your brain what to do.

When you find yourself feeling hungry, ask yourself who is feeling hungry. Scientists have traced that feeling of hunger to changes in certain hormones like leptin, but wait -- why did the leptin levels change in the first place? Could it be that a microbes someplace was manipulating your leptin levels, perhaps by poking that vagus nerve just the right way -- to get your brain to start thinking about whatever food that microbe wants?

This isn't as ridiculous as it sounds, the idea that microbes could influence your feelings and desires. Think about a disease like the rabies microbe. Because it spreads through saliva, it can't find new territory unless its host somehow finds itself exchanging saliva -- biting -- another potential host. So guess what a rabies victim can't stop thinking about? Biting a new victim. The microbe literally puts a thought into the mind.

There are many other examples, so many in fact as to be potentially a bit disturbing when we realize that we humans may be much more at the mercy of tiny microbes than we think. Links have been made between microbes and schizophrenia, stress, anxiety, self-grooming, and much more. Autism Spectral Disorders, which have always seemed puzzling because of the relationship they seem to have with digestive problems, are also linked to microbes, or the lack thereof.

Perhaps the most intriguing example is the common parasite *Toxoplasma gondii*, the strange organism that can only reproduce in the intestines of cats. A parasite seen often in all warm-blooded mammals, it's found in about a third of the global population of humans too. It's one of the reasons they tell pregnant women to stay away from cat litter. But here's the interesting part: when a Toxo protozoa infects a mouse, it leaves cysts in the mouse brain that make it

attracted to cat urine! Yes, it changes the neurology of a mouse so that it's more likely to end up inside a cat's tummy -- exactly where it can reproduce.

Think about this too much, and you'll end up with the obvious question: what other weird microbes are infecting us right now? Can we explain some of our own behaviors this way? Is there a human equivalent of these infections, driving us to do things we "ordinarily" wouldn't do? And maybe these microbes are so ubiquitous, teeming all over us and in our brains, maybe there's no way to even know what "ordinary" or "normal" human behavior is.

### **What is health?**

Western medicine tends to think reductively about health, dividing the body into pieces like organs and cells and prescribing interventions that target one particular aspect of whole, with specific drugs or supplements. But of course nothing as complex as the body and health can be simplified this much. Maybe you can't really think about human hardware without thinking about the whole ecology that goes with it, the various organisms live in and around us and greatly outnumber us.

When you think this way, then suddenly "health" takes on a whole new meaning, because we're no longer talking about the status of a single organism -- me -- but rather about the entire function "eco-system" of many, many living things, including the "me" that I want to refer to as a human.

***Microbes everywhere!***

***In the air***

Every human body is home to a rich environment of tiny living organisms, trillions of them, with their own genes, their own ways of living and surviving, and their own long histories, permanently intertwined with humanity since the beginning. These microbes inhabit every possible niche inside and outside the body. Covering every inch of skin, lining every part of the nose and ears, they have burrowed deep inside nearly every organ, consuming as well as producing, living in such quantities that together they would weigh somewhere between three and five pounds — about as much as the brain.

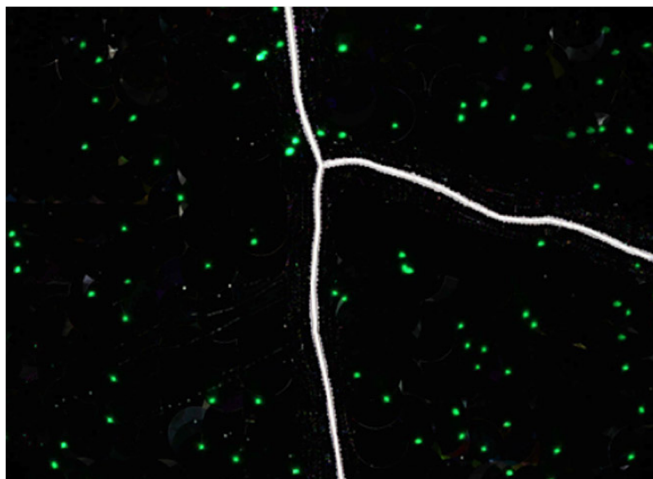
The word “microbe” refers to any tiny organism that carries its own genetic information for purposes of propagating itself. Far too small to see with the naked eye, dozens could fit inside a typical human cell. The vast majority of microbes are bacteria, though viruses and phages fit this description too, as well as archaea, the extremophile life forms that live and thrive in environments too challenging for bacteria.

A typical human will have somewhere between [n and m] [source?

] individual cells — blood, skin, neurons, other tissues — each descended from a single germ cell, the fertilized egg that began dividing itself and multiplying after the moment of conception. By contrast, microbes had to hitch their own rides later, beginning in the womb and turning into a flood at the moment of birth and beyond. But each microbe, like each human cell, remains deeply interconnected with the fates of all the other cells, all the way to the moment of death, and even after that.

[Examples of odd microbes everywhere]

A team in France, performing experiments on snowflakes, discovered microbes in the sky.



!fig(amatoSkyMicrobe)

These microbes are so numerous that they affect the climate.

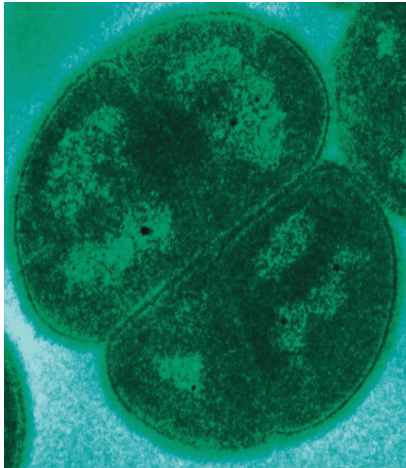
Traces of plankton have been found in space, on the surface of the International Space Station, where they are believed to have floated from the upper atmosphere.

<http://tass.ru/en/non-political/745635>

## *Underground*

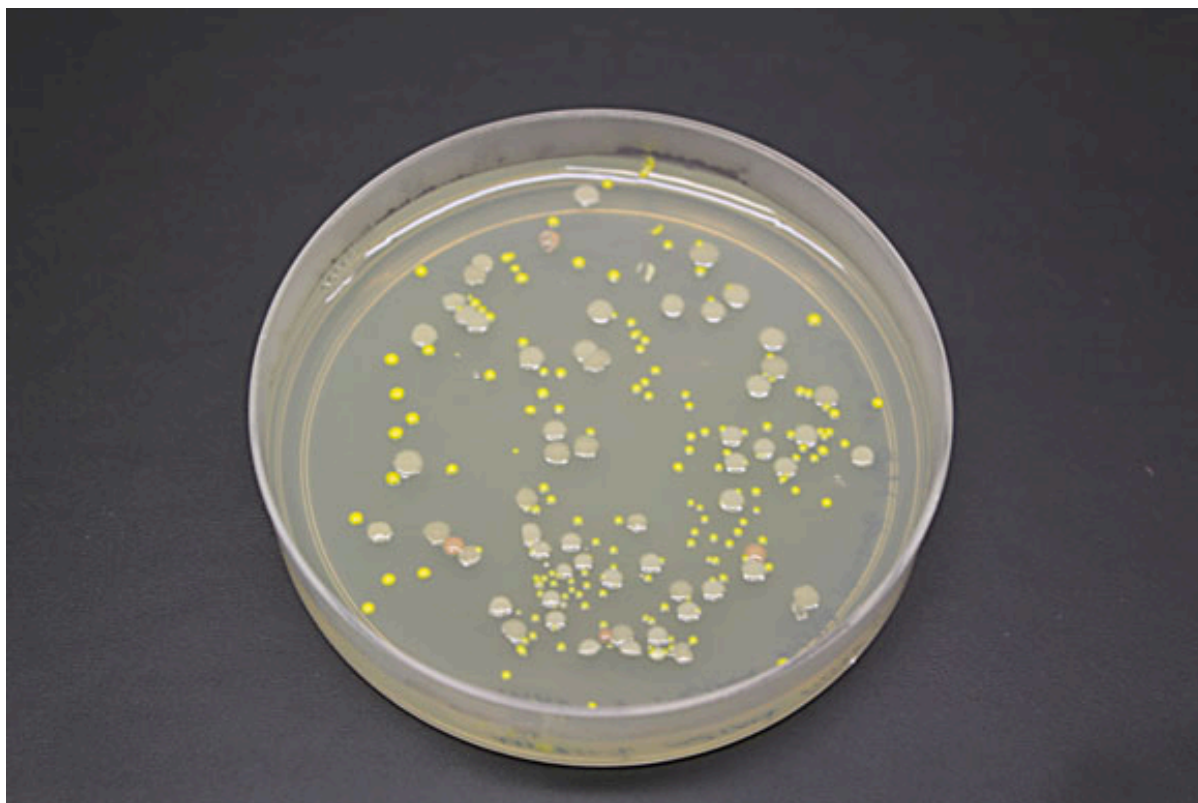
Three kilometers underground, in a South African gold mine, scientists discovered [microbe that lives on dirt]

There are even microbes that thrive in radioactive environments, like the dangerous interior of the Chernobyl nuclear reactor.



Subglacial Lake Whillans is a lake found under more than 800 meters of ice in the West Antarctic. A careful underground bore hole was inserted by a team from Louisiana State University in 2014 [<http://www.nature.com/nature/journal/v512/n7514/full/nature13667.html>] and found more than 4,000 different species of bacteria surviving under the ice. The total bacterial count was not that different from the count you'd find in surface lakes on other parts of the planet, a fact that is especially surprising for an environment that hasn't had a ray of light in millions of years. The bacteria instead thrive on iron, sulphur, and nitrogen as energy sources.

<http://earthsky.org/earth/diverse-microbes-found-deep-beneath-antarctic-ice-sheet>



***Microbes***



## ***What is a microbe?***

Scientists have known since Pasteur times that our bodies and environment are awash in other species, microbial bystanders that seem to grow everywhere. But the techniques for uncovering which organisms are where and what they are doing was revolutionized in the first decade of the 2000s by those new-fangled gene sequencers that were so usefully applied to human genes.

The old technique for studying the microbial life around us required taking a sample, inserting it into a culture of some nutrient broth known to be good for breeding microbes, and waiting a few days to see what grows (or doesn't). That's still a common way to study microbes, and that couple-of-day incubation period is one reason you don't get your lab tests back for a few days.

Now, thanks to new machines originally developed for mass DNA sequencing, the process of finding and understanding microbes has been revolutionized. It's now possible to search for new life forms *without* growing them in a culture, and this has made possible a major shift in how to think about life—and what is important and special about human hardware.

Unlike the genetic discoveries you can make by understanding your DNA (from a low-cost consumer service like 23andme), much of the news from the microbial world is *actionable*. There's little, if anything, you can do if you find you have a particular type of gene that gives you, say, a propensity to alzheimers for example. But because the microbes around you are constantly changing anyway, and because you can influence which ones grow and which don't, the world of the human micro biome is eminently *actionable*.

### **What is a “microbe”?**

Answer: it's a microorganism, the focus of the field of microbiology, and including everything from bacteria to single-

A gene is an instruction for making a protein. There are roughly 30,000 genes in your body, so there should be at least 30,000 proteins floating about, though actually it's much more complicated. The gene represents the *instructions*, but which proteins are expressed depends on a complex interaction among the rest of your DNA and its environment. The science of figuring it all out is called *proteomics*, and it's difficult to say exactly how many proteins are produced in your body.

One reason it's difficult is that *you* don't produce all the proteins. In fact, *you* carry many more genes than the 30,000 that we find from sequencing the human genome. The other genes—roughly ten times as many as those carried by your DNA—are from various other microbial life forms that are found all over, and often inside your body. The collection of all the genes expressed by these tiny microbes is called the *micro biome*, and it's enormous.

Most microbes are bacteria, and for our purposes you can assume that's what we mean, though in reality there are others: archaea (the most ancient form of

life, similar to bacteria and generally found in extremely hostile environments), viruses (which are tricky to describe as “life” because they don’t reproduce without help from other cells), and the viruses that plague bacteria (called *phages*). Each of these has a role to play in the kinds of proteins that appear in your body, though precisely which one does what is a very complicated question that science is only beginning to explore systematically.

In the olden days, before the development of gene sequencing technology, scientists studying these tiny non-human lifeforms within us had to take a sample from your body (skin, feces, nose, genitals), place it into a petri dish with some kind of broth known to be good for bacterial growth, and wait a few days to see what shows up. Bacteria reproduce asexually, through mitosis cell division, and when the conditions are right they can quickly go from just a few organisms to enough to be seen by the visible eye — sometimes in just a few hours. Nevertheless this traditional way of studying the micro biome was very time-consuming and often hit-or-miss: although *many* bacteria reproduce well outside the body, many — perhaps *most* — don’t. And even if you are fortunate enough — or clever and persistent enough — to grow a significant number of bacteria this way, there are limits to what you can find from simply observing them. It’s difficult to uncover the relationships (taxonomy) among different species, and often you’re forced to rely simply on judgement calls based on the shape, size, or other physical characteristics of the organism.

Gene sequencing technology changed all of that, especially since the middle of the first decade of the 21<sup>st</sup> century, when sequencing machines became much faster and lower cost. Instead of trying to *grow* each bacterium into a large enough colony to study, scientists can now simply scoop up an entire field of goop — from the skin, the feces, nose, etc. — And then drop the entire thing into a gene sequencer to see at a glance *all* of the genes in the sample. It’s this technology that enabled scientists to see how the genes of the micro biome had been vastly undercounted. Even if we can’t directly see the specific organisms that are making these genes, we can see the genes themselves, and then through clever software we can reconstruct the DNA for each organism.

The result is a lot of genes, as we said, at least 10x the number harbored by the DNA in each of your body cells. Now, those genes aren’t necessarily *all* making proteins that affect the body. In fact, many — perhaps *most* — of the genes are expressing something needed for the bacteria themselves: for example, instructions for how to make its cell wall, or how to reproduce. Antibiotics often target one of these bacteria-only processes in order to “gum up the works” of the bacterial life cycle without directly affecting human genes, whose own processes are governed through a whole different set of genes. Nevertheless, a great many of these microbial genes *do* play a role in the body, often a major one, and these are the ones that interest us here.

## ***Microbes in the gut***

### The Appendix

Somewhere near the border between the small intestine and the large intestine (colon), there is a small, pinkie-like appendage bulging inconspicuously into the abdominal cavity. For most of the period of modern western medicine, nobody knew its function, other as a source of severe pain when it occasionally exploded in infections that until the development of antibiotics were usually fatal.

Fun fact: the only death among the hundred or so frontiersmen and explorers on the Lewis and Clark expedition in the early 1800s was caused by appendicitis — a terrible, painful death in the wilderness but one that would have been just as inevitable at the finest hospitals in the world at the time.

What's known is that the appendix apparently houses three things: tissue from the immune system, a bunch of what are called IgA antibodies that fight infections, and tons of bacteria. What's also known is that people who have their appendix removed surgically — usually as a result of one of those terrible infections — recover to live apparently normal lives, with no side effects even decades upon decades later. So what is the appendix doing?

One clue comes from a recent study of 254 patients recovering from a *C. Difficil* infection, a horridly tenacious bacterium that has gained near-complete resistance to antibiotics and is notoriously difficult to treat. In the study, those with an appendix saw their infections come back 11% of the time, but those with no appendix suffered re-infection rates of 44% — a large and significant difference that has caused most scientists to speculate that the appendix is harboring something that enables the body to recover after a microbial disaster.

My own interest in this subject is not idle curiosity. About fifty years ago, doctors at a small town hospital gave an appendectomy to my five-year-old self in order to treat an unusual belly ache. While doing what was apparently an exploratory surgery to find the cause, they came upon my otherwise healthy-looking appendix and decided on the spot to simply remove it. “Why not?” I’m sure they thought at the time. “If nothing else, it’ll prevent him from appendicitis”, which is as tautologically sound a reason as any.

## ***Microbes of the skin***

### Microbes of the skin

Microbes of the gut are important, but many other organisms are crawling all over you too.

It should be no surprise that the most obvious place for them to live is on your outside: your skin and hair. We usually think of bacteria, fungi, or viruses, as something to eradicate, pests that can only be there to cause problems. The natural state of the body is pristinely washed: clean, with no dirt or other blemishes.

This pristine state has never been true. Humans, like all other multi-celled creatures, have lived forever in a bath of foreign ride-hitchers of all types. This micro-fauna is as natural to us as the air we breath, and like air, our bodies *need* a good environment to be at our best.

The body maintains its temperature well, occasionally cooling itself off by channeling excess water through the sweat glands and onto the surface of the skin. Besides water, this sweat contains small amounts of other matter, including elements that toss molecules into the air, many of which, upon landing on the olfactory glands of another person will be sensed as a smell — an odor, body odor.

Nature does not like waste, especially not anything associated with a living body like us, and as soon as those micro elements in sweat are exposed to the air, in the natural state there will be other lifeforms available to devour them, taking out some of the leftover energy in the process of converting them to simpler chemical compounds like carbon dioxide.

One such hanger-on, is the bacterium *Nitrosomonas eutropha*. It thrives in your skin and makes a living from oxidizing, that is, converting to oxygen, the ammonia that is naturally present in the waste secreted by your sweat glands.

### ***A crash course in microbiology***

Before we learn how to study your own microbiome, let's review some of the science.

All life runs on three chemical building blocks: DNA, RNA, and proteins. Each of these is an arbitrarily-long chain of repeating molecules called *nucleotides* (DNA or RNA) and *amino acids* (proteins). Due to constraints on the way atoms interact, the set of building blocks is fixed. All DNA is composed of only four nucleotides: adenine, thymine, guanine, and cytosine, represented by the letters ATGC. RNA is composed of the same molecules, except that uracile (U) is substituted for thymine.

Similarly, proteins are constructed with only 20 different amino acids, which can again be represented by a short three-character abbreviation.

The correspondence between these different proteins and combinations of DNA or RNA is referred to as the genetic code.

nonpolar    polar    basic    acidic    (stop codon)

#### **Standard genetic code**

1 s t b a s e	2nd base								3 r d b a s e
	T		C		A		G		
T	T T T	(Phe/F) <u>Phenylala nine</u>	T C T	(Ser/S) <u>Serine</u>	T A T	(Tyr/Y) <u>Tyrosine</u>	T G T	(Cys/C) <u>Cysteine</u>	T
	T T C		T C C		T A C		T G C		C
	T T A	(Leu/L) <u>Leucine</u>	T C A		T A A	<u>S t o p</u> (Ochre)	T G A	<u>S t o p</u> (Opal)	A
	T T G		T C G		T A G	<u>S t o p</u> (Amber)	T G G	(Trp/W) <u>Tryptopha n</u>	G

C	CTT		CCT	(Pro/P) <u>Proline</u>	CAT	(His/H) <u>Histidine</u>	CGT	(Arg/R) <u>Arginine</u>	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	CGA	A		
	CTG		CCG		CAG	CGG	G		

A	A		A		A		A		
	T		C		A		G		T
	T		T		T (Asn/N)		T (Ser/S)		
	A	(Ile/I)	A		<a href="#">Asparagine</a>		A <a href="#">Serine</a>		
	T	<a href="#">Isoleucine</a>	C		A		G		C
	C		C (Thr/T)		C		C		
	A		A <a href="#">Threonine</a>		A		A		
	T		C		A		G		A
	A		A		A		A		
	A				(Lys/K)		(Arg/R)		
G	T	(Met/M)	A		A <a href="#">Lysine</a>		A <a href="#">Arginine</a>		
	G	<a href="#">Methionine</a>	C		A		G		G
	A		G		G		G		
	G		G		G		G		
	T		C		A		G		T
	T		T		T (Asp/D)		T		
	G		G		<a href="#">Aspartic acid</a>		G		
	T		C		A		G		C
	C	(Val/V)	C (Ala/A)		C		C (Gly/G)		
	G	<a href="#">Valine</a>	G <a href="#">Alanine</a>		G		G <a href="#">Glycine</a>		
	T		C		A		G		A
	A		A		A (Glu/E)		A		
	G		G		<a href="#">Glutamic acid</a>		G		
	T		C		A		G		G
	G		G		G		G		

<sup>A</sup> The codon ATG both codes for methionine and serves as an initiation site: the first ATG in an [mRNA](#)'s coding region is where translation into protein begins.<sup>[1]</sup>

**Inverse table (compressed using [IUPAC notation](#))**

Amino acid	Codons	Compressed	Amino acid	Codons	Compressed
Ala/A	GCT, GCC, GCA, GCG	GCN	Leu/L	TTA, TTG, CTT, CTC, CTA, CTG	Y T R , CTN
Arg/R	CGT, CGC, CGA, CGG, AGA, AGG	C G N , MGR	Lys/K	AAA, AAG	AAR
Asn/N	AAT, AAC	AA Y	Met/M	ATG	

<b>Asp/ D</b>	GAT, GAC	GAY	<b>Phe/ F</b>	TTT, TTC	TTY
<b>Cys/ C</b>	TGT, TGC	TGY	<b>Pro/P</b>	CCT, CCC, CCA, CCG	CCN
<b>Gln/ Q</b>	CAA, CAG	CAR	<b>Ser/S</b>	TCT, TCC, TCA, TCG, AGT, AGC	T C N , AGY
<b>Glu/ E</b>	GAA, GAG	GAR	<b>Thr/T</b>	ACT, ACC, ACA, ACG	ACN
<b>Gly/ G</b>	GGT, GGC, GGA, GGG	GGN	<b>Trp/ W</b>	TGG	
<b>His/ H</b>	CAT, CAC	CAY	<b>Tyr/Y</b>	TAT, TAC	TAY
<b>Ile/I</b>	ATT, ATC, ATA	ATH	<b>Val/ V</b>	GTT, GTC, GTA, GTG	GTN
<b>STAR T</b>	ATG		<b>STOP</b>	TAA, TGA, TAG	T A R , TRA

As a programmer looking through all of this, you may immediately be inspired to write your own software version of this. After all, the remarkable consistency between all of these building blocks cries out for manipulation by computer.

In fact, that's exactly what bioinformaticians do, and numerous software packages have been developed to make it easy to treat these building blocks of life like ordinary computer strings.



Perhaps the biggest challenge is the volume of data to be handled, which can easily be measured in gigabytes for a simple organism, but can require entire server farms in the case of some real-world biological systems. For that reason, much of bioinformatics is about optimizations to improve the speed of processing a large data set, or to simplify the presentation in a way that can reveal the most biologically interesting aspects of a problem without wading in over-complexity.

One special protein, DNA, can store information.

## **The cell**

Start with the cell. All living things are composed of cells, which are tiny self-contained blocks where entropy is kept at bay for as long as possible. Everything in the universe tends over time to fall into disorganized entropy, but cells contain many tricks, honed over billions of years of evolution, to thrive.

Some organisms have a single cell, and others are multi-cell, but the key division of life is not between the number of cells but rather in the complexity of its internals. Human cells, like all multicellular life forms, are classified as *eukaryotes*, because all of our cells are made of a number of internal features, called *organelles*, of which the most prominent is the *nucleus*.

Bacteria, by contrast, are *prokaryotes*. Their cells look much simpler under a microscope: just a cell wall bounding some free-floating objects.

Part of the definition of life is the ability to reproduce, and cells do this by dividing themselves in half in a process called *mitosis*.

## ***A crash course in botany***

How do you talk about the relationships between various different life forms?

A taxon is a simple unit of life. A *homo sapiens* is a taxon, but so is a primate. A mammal is a taxon too. It might seem odd in the ordinary biological world to bother using the same term 'taxon' to refer to all of those units, but for bacteria and anything that reproduces asexually, it's an important distinction because often, taxonomists don't agree about whether a group of organisms is part of the same taxon or not.

Since Carl Linnaeus in the 1700's, the science of taxonomy divides all life into seven major categories: Kingdom, Phylum, Class, Order, Family, Genus, Species (which I was taught in sixth grade to remember by the mnemonic "King Philip Came Over for Girl Scouts").

Bacteria make up their own kingdom. Just as the animal kingdom includes everything from humans to jellyfish to beetles, the diversity of bacterial life is enormous, a point which can't be emphasized too much. This is true at every rank in the taxonomy. Even two organisms that are the same at a lower rank, like genus, might have radically different effects on the human body, just as a member of the animal genus *Canis* could be anything from a wolf or coyote to a Chihuahua.

You cannot mix and match these ranks. If you know something about the number of organisms in one genus, for example, this is meaningful only in comparison to the numbers of another genus. Keep that in mind during our analysis.

### ***What are they doing?***

*Toxoplasma Gondii* is a tiny microbe that, for some reason, only likes to reproduce from within the gut of a cat. It can be found in almost all warm-blooded mammals, including humans — about 30% of us, according to some estimates, and that's after a century of obsession with hygiene that has wiped out countless other tiny inhabitants of the body.

*T. Gondii* seem harmless because its hosts appear to show no differences before or after “infection” except in one creature: the mouse. When it finds itself ingested by a mouse, it appears to fade away quietly. MRI scans show large amounts in the gut for a week or two, gradually decreasing until there is apparently nothing. Except: tiny *T. Gondii* cysts begin to show up in the mouse's brain. Not everywhere, but in just a few strategic places.

Then something strange happens: the mouse is suddenly attracted to cat urine. And note: its only attraction is to *cat* urine, not rabbits, not humans — only cats. Everything else about the mouse appears normal. It still seems to be afraid of the other things that scare mice: other predators, stressful situations. But put a *T. Gondii*-affected mouse into a maze with different animal urine in the corners and it will rush to the cat side every time.

How is it that *T. Gondii* is able to be so precise in its effects? It only seems to affect mice, and even then it only apparently makes them attracted to cats. It doesn't apparently cause any other harm to its hosts, so — somehow — it must have found a resting place right at the spot in the mouse brain that affects its interest in cats.

Although there doesn't seem to be any major negative consequences to humans hosting *T. Gondii*, some scientists aren't sure. One man in particular, Jaroslav Flegr, an evolutionary biologist at Charles University in Prague, thinks he has evidence that women who carry it might be more trusting than those who don't.

*T. Gondii* isn't the only microbe known to affect the behavior of its host. A more common example is the rabies virus, which upon infection somehow causes a mammal to be more agitated, more likely to strike out — or bite — other humans, thereby spreading itself.

Or syphilis, the disease spread by the bacterium [*sypilius*] and causes its host to go insane. The microbe is somehow able to infect the mind of the victim.

In general, sexually transmitted diseases are especially likely to have behavioral consequences. A sexual disease that produces symptoms is unlikely to spread, yet it still requires contact with a new victim. Perhaps the ideal vector is a behavioral change, making the host more likely to come in contact with a new host.

Think about this too much, and you'll end up with the obvious question: what other weird microbes are infecting us right now? Can we explain some of our own behaviors this way? Is there a human equivalent of these infections, driving us to do things we "ordinarily" wouldn't do? And maybe these microbes are so ubiquitous, teeming all over us and in our brains, maybe there's no way to even know what "ordinary" or "normal" human behavior is.

## ***Studying the Human Microbiome***

### Studying the Human Microbiome

During the late 1990s, biologists around the world were racing to complete the first sequence of the human genome, a first draft of which was available in June 2000, and the project was declared complete in 2003. With this sequence available, scientists now had the first rough map of all the genes that make up a human being, and they were hopeful that this would be quickly followed by new discoveries linking genes to health.

The Human Microbiome Project was launched in 2008 by the National Institutes of Health with a budget of \$115M.

## CHAPTER TWO

### *The Technology*

## ***Studying the Microbiome***

Two major problems had to be solved before scientists could begin seriously studying the microbiome.

First, the various bacteria inside us had to be identified. Since the invention of the microscope, the process was straightforward: collect enough of the organisms of interest and then look at them, either as a whole while cultured in a lab, or individually under the microscope.

Computer scientists have an easier time than biologists because operations are so fast you can afford to do them over and over. In biology, individual operations are much slower, but you can run them in parallel. That goop in the petri dish is often made of trillions of organisms, each one consuming, producing, and multiplying individually. Even with the very low error rates in ordinary cell division, you have so many organisms that you will regularly get a few mutations, and if the experiment is well-controlled, you will be able to see those mutations.

### ***The old way: culturing***

How do you tell which microbe is which?

Until recently, the old fashioned way was the only way. You get some of the bacterium of interest and then grow more of it. In other words, a glorified version of farming.

How do you farm a microbe? Basically, you put it in a petri dish along with something that it likes to eat, and you wait for it to grow. What does it like to eat? Well, that's the "art". If you find the right substance and just the right environment, it might grow. Then again, it might not. Much of this is trial and error.

A bigger problem with culturing is that some organisms need other organisms in order to survive. You may be simply unable to grow the one of interest unless you have a whole colony of other, unrelated but symbiont organisms nearby to provide it something it needs. In other words, it may simply be impossible to grow a bunch of one particular organism in isolation.



## 16s rRNA

### The 16S shortcut

A living organism can contain somewhere between thousands and hundreds of billions of DNA base pairs. There is some commonality between related organisms — humans and chimpanzees, for example, share upwards of 90% of their DNA — but in general it's hard to use the DNA strand itself to measure the relatedness between two organisms. Understanding the reason for this may help you understand why there is a clever shortcut.

You might think you can measure the relatedness of two organisms by looking at all the DNA in each one, and computing the percentage that each shares in common. This would work, but sequencing all those billions of DNA bases takes a lot of time and money, and it would be impractical in a case like the microbiome where you may need to do this for millions of individual organisms.

A service like 23andme is able to cheaply compare individuals of the same species (i.e. Humans) because the generic human genome is already well-mapped and we know that of the 3 billion base pairs, only about 3 *million* (the SNPs or single-nucleotide polymorphisms) are different between individuals. When you give your spit sample to 23andme, they give you back a subset of your SNPs, only those that have been studied enough to be interesting. SNPs are easy and cheap to find using a “gene-chip”, a special semiconductor-like device that can quickly look at 1 million or more pre-determined spots on your DNA. But this is only possible because the map itself already exists, thanks to multi-year effort of the Human Genome Project that finished in the early 2000s. There are no gene chips (yet) for bacteria, and certainly not for all the millions of types in your body. And even if there *were* such chips, bacteria are notorious at adapting and changing to their surrounding environment, exchanging genes with one another, that it just wouldn't be practical to identify enough constant genes to make it worthwhile.

Fortunately, some parts of an organisms DNA change more often than others. The SNPs are reliably different between any two individuals of the same species, but some of parts of DNA are so important that they stay virtually identical even across entire families of organisms. Something like skin pigmentation, for example, can change without having much effect at all on survival, but remember that DNA describes absolutely *everything* about the organism, including the workings of very low-level cell process. Not just the skin pigmentation, or even how skin cells grow, but much more fundamental: how a cell divides, for example, or even how to use the oxygen a cell needs for survival.

Among the most fundamental of all processes is the what happens in every cell's ribosome, a special enzyme that knows how to copy DNA. The ribosome is such a fundamental part of every living organism that very little about the ribosome has changed in a billion years of evolution. Humans and corn plants

actually share quite a bit of the ribosome; both are prokaryotes, for one thing, so many of our cellular processes work the same.

But bacteria go back even further than humans and corn plants, enough so that the differences aren't so subtle anymore. In fact, the differences are big enough that you don't even need to do gene sequencing to peek at one of the most

If you cut up bacterial cells ("lyse" them, to use the technical term) to pull out their DNA, and you throw it into a centrifuge turning at a carefully controlled speed, some the "goo" from the cells will rise further than others because some parts of the cell are heavier. One specific part of the cell, corresponding to a section of the cell's ribosomes, will rise to a centrifuge level referred to as "16S". Precisely skimming the goo at that spot will give you a collection of RNA from just one part of the ribosome of bacteria. Other cell types that may happen to be in the sample will not show up at that spot on the centrifuge. That's the shortcut.

Once you have a bunch of that 16S ribosomal RNA, you know that you have bacterial RNA and *that's* what the Illumina NextSeq 500 will sequence. This represents a huge subset, a nearly trivial percentage of all the DNA in the overall sample, but combined with one more shortcut, it's enough to accurately identify the sample.

The remaining shortcut is possible thanks to years of research of sequencing the genes in bacteria. Scientists in labs around the world have been faithfully digging up samples of bacteria, and performing whole-gene sequencing on what they find. These sequences are kept in a large public database run by the National Institutes of Health, and if you have a snippet of DNA — even a small snippet — you can do a string compare against all the genes that have already been analyzed. When you find a match, you can be pretty sure that your sample contains this type of bacteria.

It's this two-step combination, 16S "skimming" and a database lookup, that makes it cost effective to study the millions of organisms in your microbiome. You don't have to do a complete gene sequence on every single bacterium; just trust that the tiny subset of RNA in the 16S region is enough to uniquely match something already in the bacterial database.

## ***Gene sequencing***

Gene sequencing

## ***Bioinformatics***

Once you have the data computerized, the analysis tools leave the wet world of biology and enter the software world of computing. This marriage is called bioinformatics.

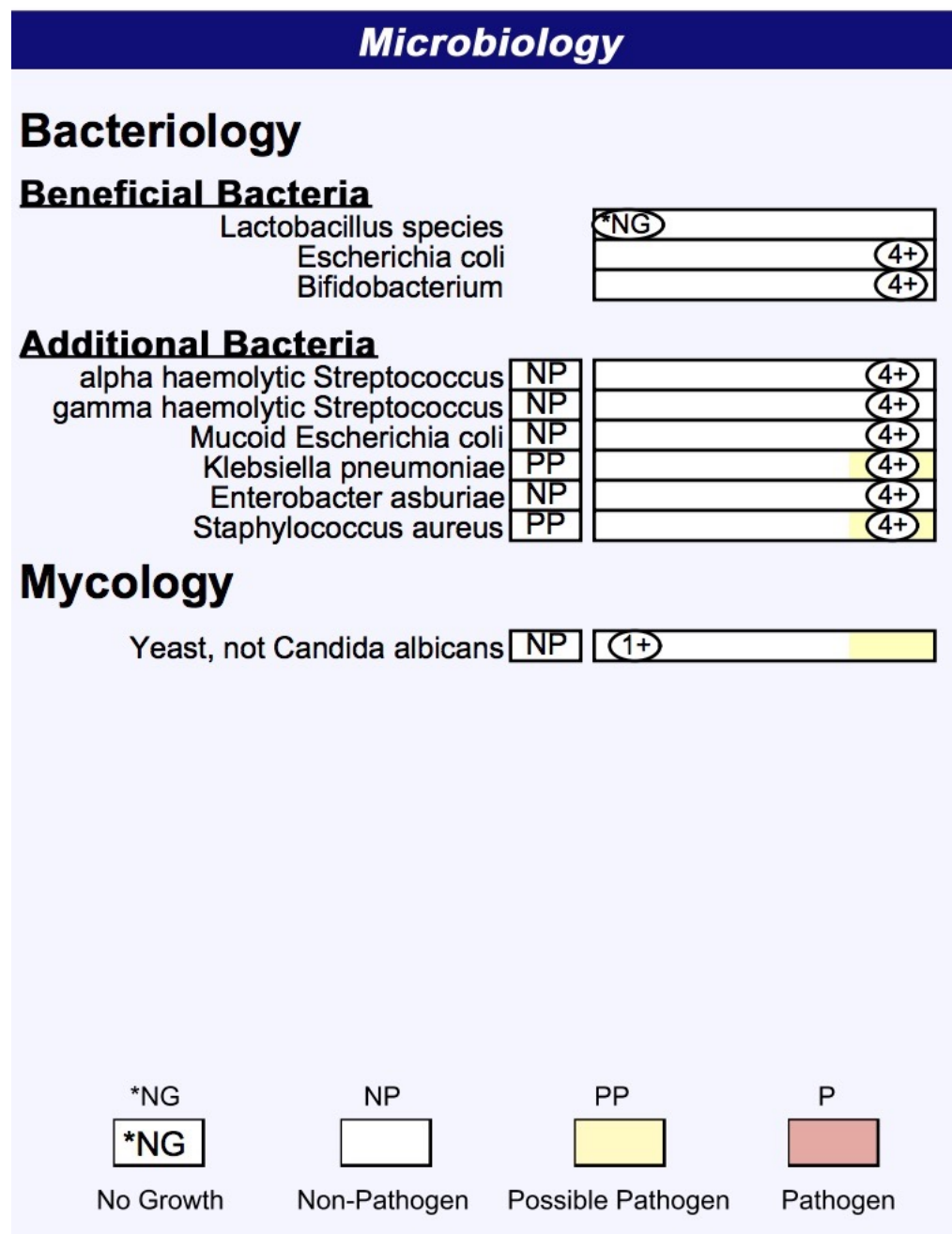
## **CHAPTER THREE**

### ***How to Analyze Your Microbiome***

## How to get your data

There are several ways for an individual to get access to the powerful sequencers that make 16S rRNA analysis possible.

One way is through a doctor, or other health care provider who can give you a clinical test, just like you might do a blood draw as part of an annual physical. If you have a medical condition, the tests are likely covered by insurance, but tend to be pretty expensive (several hundred dollars or more). Genova Diagnostics is the market leader, and will give your doctor a report that looks something like this:



Though expensive, Genova results are intended to be clinically significant. Many of their diagnostics are FDA approved, so your doctor can be assured

they're of the highest quality. They test beyond the microbiome, for example, looking at organic acids that are often associated with gut problems. They will also identify parasites, which won't be detected with 16S technology.

On the other hand, since Genova tests are intended for a doctor, the best you can hope for is a PDF copy of their final results, similar to the chart above. If you're an experimenter, this won't be nearly enough, and you'll soon want the raw data.

For "normal" people, just curious about the science like I am, there are two other ways to find out your microbiome:

The first is called American Gut Project, associated with a lab at the University of California San Diego and others. For a \$100 donation, they'll send you a sampling kit which you send back to get a breakdown of your gut bacteria. American Gut boasts an impressive number of practicing research scientists, including several of the key contributors to the Human Microbiome Project.



I recommend the San Francisco company uBiome.

### **History of uBiome**

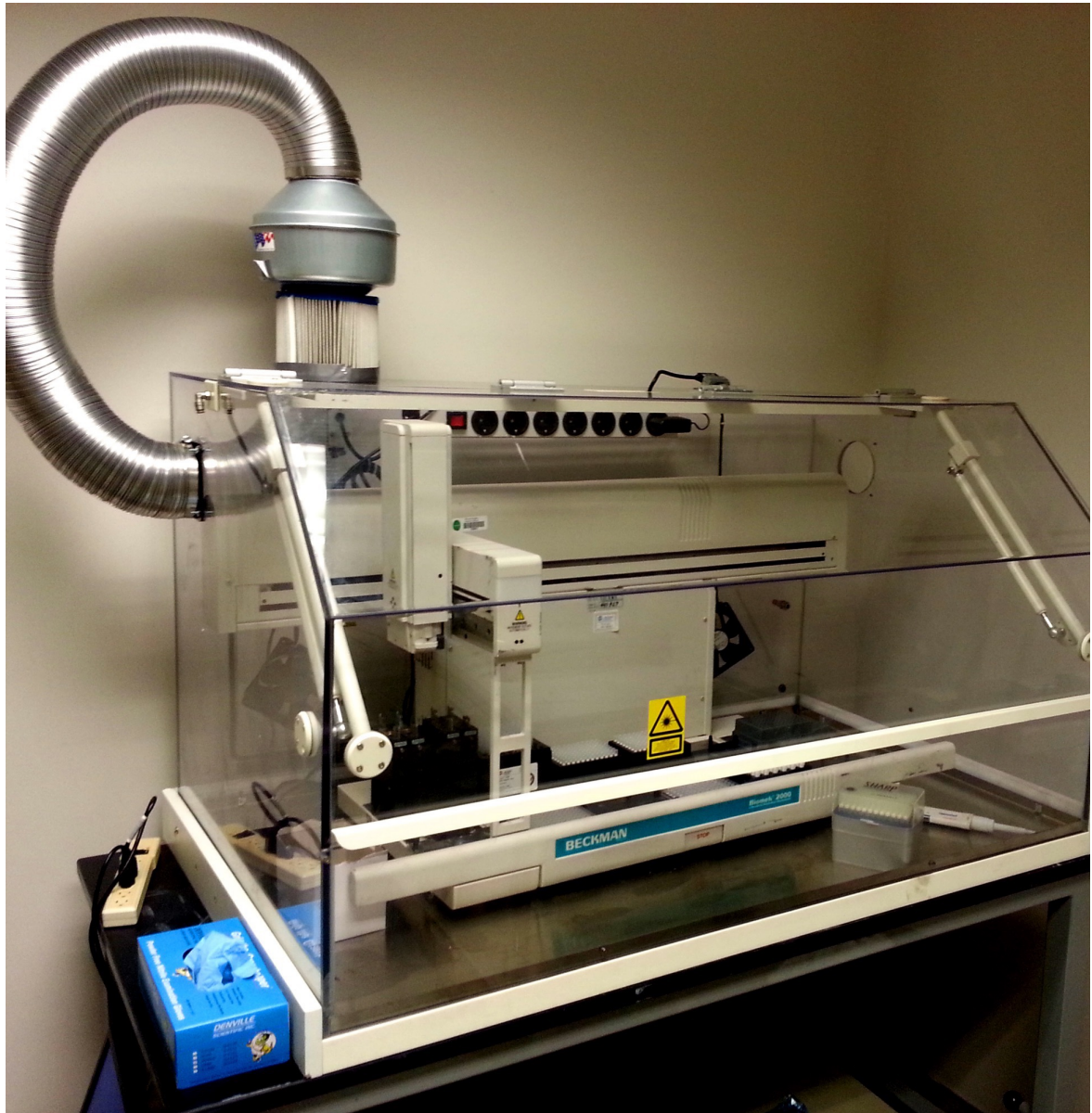
The company was founded in late 2012 by Jessica Richman and Zach Apte, who met while doing graduate studies at Oxford. Jessica, with an undergraduate degree from Stanford, was studying public policy, using computers to understand how new ideas go viral. Zach was a biologist studying the microbiome, disappointed at how long it took to collect the vast number of samples required to do the science. The problem, they realized, was not lack of interest from the general public. There was no shortage of people interested in the microbiome and its affect on health, many of whom would happily contribute samples if they could help the science. They suspected that many of these "citizen scientists" were so eager to help that they might even *pay* for the privilege of participating in an experiment, particularly if they could receive personalized results of their own tests.

In November 2012, they launched their first crowd funding campaign on the site Indiegogo, hoping to raise \$100,000 to start the business. They started by selling "perks", \$89 for a gut kit, \$399 for a five-sample kit, offering to run the samples through the sequencing machine that Zach Apte was using at the University of California San Francisco biotech startup garage. To their surprise,

the campaign raised more than \$350,000, and the company was off and running.

Since uBiome focused from the beginning on crowd-sourced open science, their kits were designed to be as easy to use as possible. With glossy-printed, step-by-step directions and an attractive kit with well-labeled test tubes, they wanted to let ordinary people take samples and submit them with as little fuss as possible.

At the lab, uBiome also focused on automation. They built two special collection robots that could open the test tubes with the ultimate goal that they'd be able to submit to the sequencing machine with barely any human intervention.



The strong start with the Indiegogo campaign brought them enough money to start the business, but more importantly it attracted the attention of other investors who soon contributed \$1.5M to the new company. Eventually the



uBiome staff signed on to the prestigious Y-Combinator, the startup incubator that has given birth to many famous names including Airbnb, Reddit, Dropbox and others, and landed an additional \$3M investment from the prestigious venture capital firm Andreessen Horowitz.

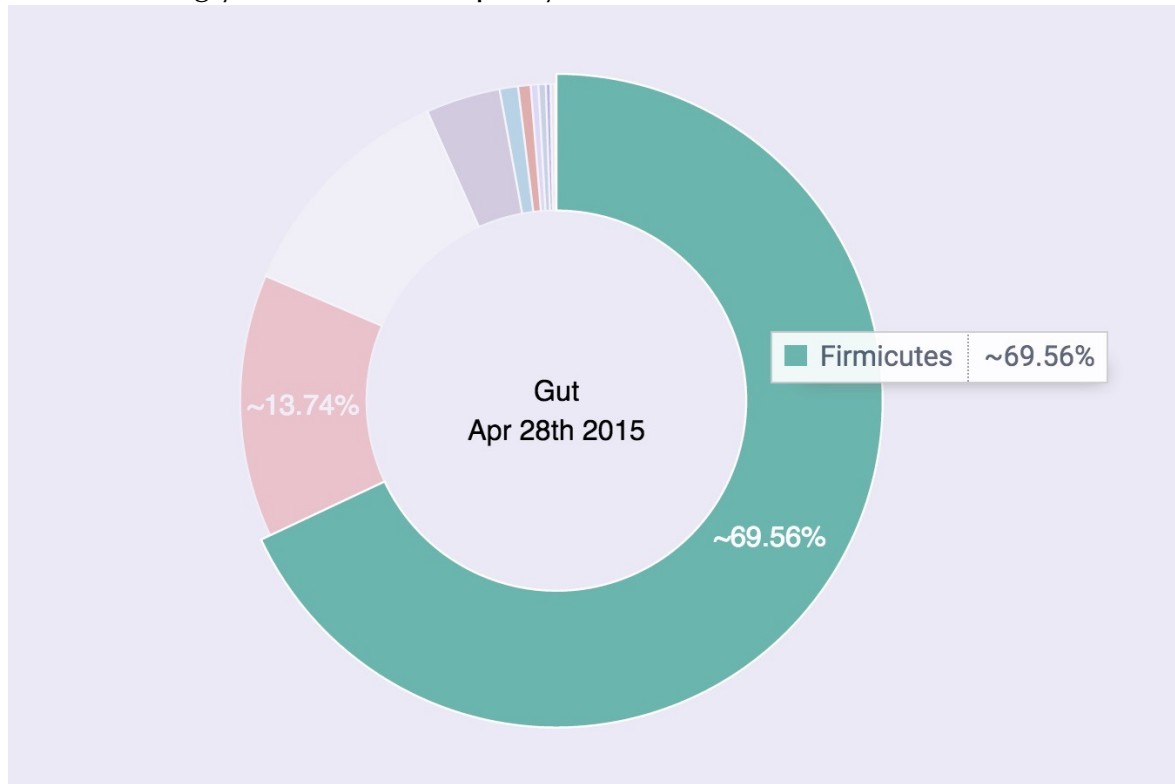
With the new investment capital, uBiome was able to open a new headquarters and buy their own \$300,000 NextSeq500 gene sequencing machine and more automation to help them process the samples more quickly.

The most popular gene sequencer is the NextSeq500, from San Diego-based Illumina.

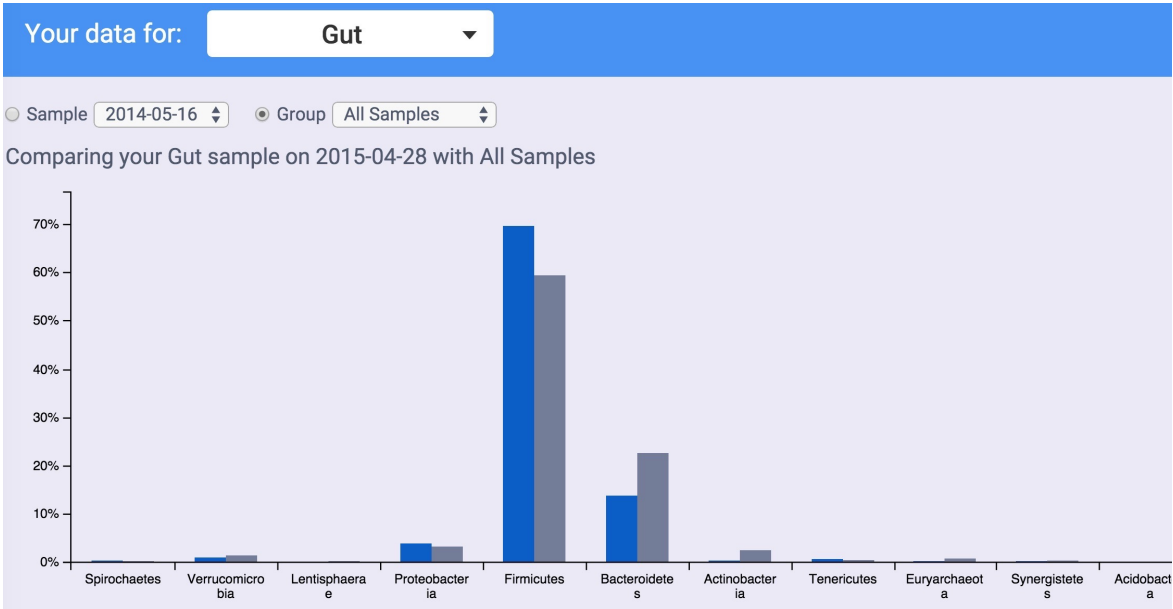
### ***The basics: web tools***

After uBiome has notified you of your results, you can go to [their web site](#) to learn more.

The first thing you'll notice is a pretty chart like this one:



This shows the overall breakdown of the bacteria they found in your gut. In this chart, the bacteria are lumped by phyla —the highest, broadest level of classification. If you're like most Americans, the vast majority of your sample will be one of two phyla: *bacteroidetes* or *firmicutes*.



## ***Analyze with Excel***

The Illumina machines spit off huge amounts of data (FASTQ files can be hundreds of thousands of lines long), so to be useful you'll need a summary. UBiome provides a much more concise version called a "raw taxonomy" file, which is generally only a few hundred lines showing just the organisms they think they found in the sample.

### **What is JSON**

UBiome's taxonomy files come in a simple structured text format called JSON (JavaScript Object Notation), commonly used across the web. Although it's not as convenient for people as the rest of the uBiome web site, programmers refer to JSON as "human readable", because if you squint enough you can sorta tell what it means without a computer program. Here's a sample:

```
{
  "ubiome_bacteriacounts": [{
    "taxon": "2",
    "parent": "131567",
    "count": "22691",
    "count_norm": "1000000",
    "avg": null,
    "tax_name": "Bacteria",
    "tax_rank": "superkingdom",
    "tax_color": null
  }, {
    "taxon": "1239",
    "parent": "2",
    "count": "15414",
    "count_norm": "679300",
    "avg": null,
    "tax_name": "Firmicutes",
    "tax_rank": "phylum",
    "tax_color": "5E6591"
  }
}
```

JSON is just structured data. That's it. The data is organized according to *keys* and *values*. The *keys* are unique identifiers and the *values* are their values. The

whole file looks that way. The only requirement of a JSON file is that it be consistent with this pattern, because after all it will need to be read by a computer eventually.

The uBiome Taxonomy JSON files identify the key `uBiome_bacteriacounts` which is mapped to a bunch of fields, each representing a particular organism found in your sample. Look at the figure above to get the basic idea.

Because uBiome JSON files are well-structured, it's easy to process them with other software, including Excel. That's our next step: bring this data into Excel.

### Converting uBiome taxonomy JSON to Excel

If you just have one or two files, it's easy to bring uBiome taxonomy data into Excel. Go to the uBiome apps page and click on the section labeled "Download taxonomy".

#### Download my sample data



**Advanced users:** click below to get your raw results to analyze your own way.

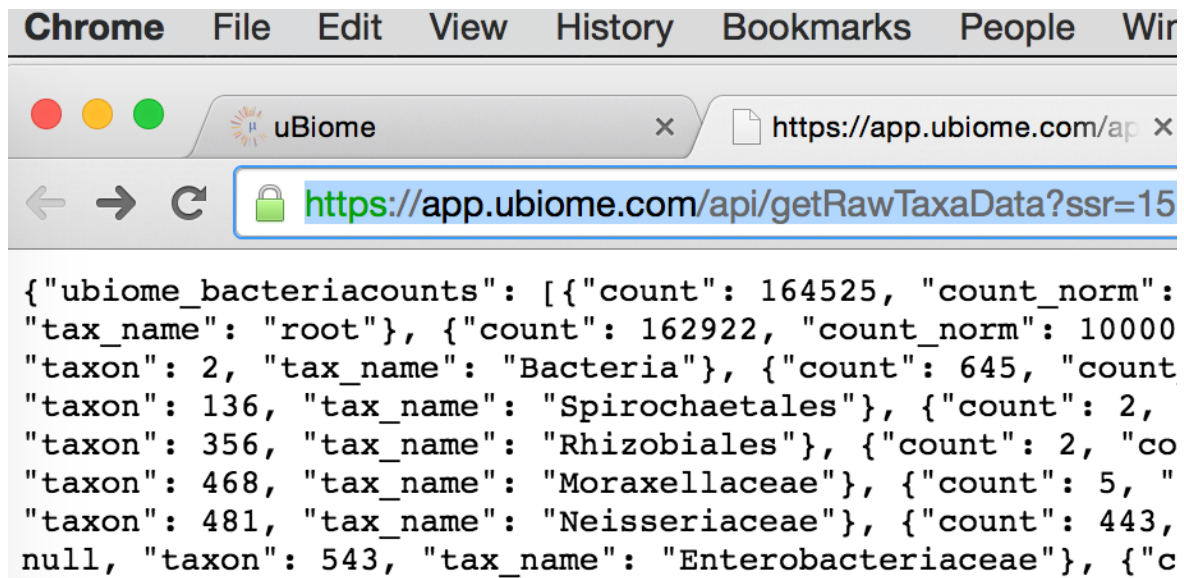


Download raw data



Download taxonomy

You'll see a page of JSON representing what uBiome found in your sample.



On a desktop computer, if you select-all (press control-A or command-A) then you can copy this data to the clipboard to prepare to paste it to another web site.

There are many web sites that will convert JSON to a format that will work in Excel. Google “Json to Excel” or “Json to CSV” and you’ll find one. Just paste your data into one of those sites to convert it for either Excel’s native (XLS or XLSX) format, or the more universal CSV format, readable by Excel. Here are a few of the sites I’ve tested:

<http://www.convertcsv.com/json-to-csv.htm>  
<http://www.json-xls.com/json2xls>

Once the data is in Excel, you can work with it just as you would with any Excel data sheet. Here’s how one of my recent uBiome taxonomy files looks in Mac Excel:

	A	B	C	D	E	F	G	H	I
1	count	count_norm	parent	tax_rank	tax_color	avg	taxon	tax_name	
2	164525	1009839	0	root	null	null	1	root	
3	162922	1000000	131567	superkingdom	null	null	2	Bacteria	
4	645	3959	203692	order	null	null	136	Spirochaetales	
5	2	12	28211	order	null	null	356	Rhizobiales	
6	2	12	72274	family	null	null	468	Moraxellaceae	
7	5	31	206351	family	null	null	481	Neisseriaceae	
8	443	2719	91347	family	null	null	543	Enterobacteriaceae	
9	3	18	543	genus	null	null	547	Enterobacter	
10	3	18	354276	species	null	null	550	Enterobacter cloacae	
11	9	55	122277	species	null	null	554	Pectobacterium carot	
12	4	25	543	genus	null	null	561	Escherichia	
13	2	12	543	genus	null	null	629	Yersinia	
14	2	12	629	species	null	null	630	Yersinia enterocolitica	
15	2	12	135625	family	null	null	712	Pasteurellaceae	

Yours may look slightly different. Sometimes uBiome changes the labels on the taxonomy slightly, and the ordering of the columns may be different depending on how the JSON was converted, but none of that matters. The key is that you are now able to work with it in a full-blown spreadsheet.

### The uBiome data tags

Let's review the meaning of each of the columns. The uBiome JSON taxonomy data includes the following fields:

The uBiome JSON taxonomy data includes the following fields:

**count:** an absolute measure of number of organisms found in the sample. Without knowing the size of the sample, or how many times the DNA inside was processed through PCR amplification, this number doesn't mean much except in relation to other counts at the same taxonomical rank.

**count\_norm:** a "normalized" version of the count, created by simply dividing the count on a given row by the count found on the row with tax\_name = Bacteria and then multiplying by one million. It's easier to just think of it as parts per million: each unit is 1 / 10,000th of a percent. For example, if you see a row with count\_norm = 500,000, you can just think of that as 50% of the sample.

**tax\_name:** this is the classification of the organism based on the level of its taxonomy. If you were looking at a human being, for example, you would see homo sapiens if you selected tax\_rank = species, but you'd see mammalia if you selected tax\_rank = class.

**tax\_rank:** tells the level of the taxonomy. In daily conversation about animals or plants, we usually refer to the species (e.g. homo sapiens), but sometimes it's more useful to talk about bigger groupings of related organisms. For example, humans are members of the class mammalia, along with tigers and horses. If this spreadsheet were counting organisms at the level of class mammalia, the count\_norm would almost certainly be bigger than the count\_norm for humans alone, unless humans were the only type of mammal found in the sample.

**taxon and parent:** these help identify the ranking in a more precise way by pointing out which tax\_ranks are subsets of which. For example, Bacteroidia above has a parent = 976, meaning that it is a subset of the taxon 976, Bacteroidetes. When you follow the various taxons and parents up the chain, you'll see they all end in the superkingdom Bacteria, which has a taxon of 2. The values for these numbers, incidentally, are taxonomical numbers from the curated database at NCBI, the national bioinformatics center run by the U.S. government. Enter the number into the taxonomy browser at the NCBI Taxonomy Browser and you can learn as much as you want about that organism.

**tax\_color** doesn't matter for this analysis, but uBiome software uses this to colorize their pretty graphs to make them more readable.

Of these fields, the most important ones are **tax\_name**, **tax\_rank**, and **count\_norm**.



## Comparing samples

The uBiome JSON taxonomy file is extremely helpful at giving you a short summary of what you care about most in a sample: the organisms within it and their relative amounts. But even with these annotations, there's a limit to what you can learn in a single sample. To be able to start hacking your microbiome, you will need to learn how to compare multiple samples.

There are two main things you'll want to measure when comparing two samples:

**Uniqueness:** which organisms are found in only one sample and not the other? If you are comparing two samples from the same individual (e.g. Yourself), then uniqueness is another way of talking about either *extinctions* (when a taxon has disappeared entirely in a later sample) and *appearances* (when a new taxon magically shows up). In other words, you care a lot about the order in which the samples were taken.

On the other hand, if you're comparing two separate individuals, then the ordering of the samples doesn't matter. Uniqueness just tells which organisms are unique to a specific person.

**Relative abundance:** which person has more of which organism? There are two senses in which we care about the relative frequency of an organism's occurrence in a sample. You might care, among all the taxons found, which ones are more highly represented in *absolute* amounts? Or you might care on a relative basis. To help understand why this matters, it may help to think of the following specific example:

In many Americans the phylum *Firmicutes* makes up a majority of the sample, often reaching 60% or more. You might find two people, one who has 60% *Firmicutes* and the other with 30% — only half as much. The relative difference is 2x and the absolute difference is 30 percentage points.

On the other hand, you might find *Bifidobacterium* makes up 10% of the first sample and 5% of the second. In this case, the relative difference is 2x — just like the *Firmicutes* case — but the *absolute* difference is only 5 percentage points, much less than in the first example.

The built-in uBiome web tools look for *relative* differences. A taxon might be a minuscule component of both samples, but when you start from a small base, even a few additional organisms can make up a big percentage change. In the extreme case, a sample with a **count\_norm** of 100 compared to another sample with a **count\_norm** of 300 will show a 3X increase — very high in aBiome terms, even though the absolute difference is only  $300 - 100 = 200$  organisms.

Compare that with a an organism with a **count\_norm** of 50,000 in one sample and 51,000 in another sample. UBiome's algorithms will treat these as a

virtual tie, even though the absolute number of organisms in the second sample outnumber the first by 1,000 organisms.

Which method of measuring is better? It depend on what you care more about. Some species need have only a tiny representation in a sample to make a big difference in health outcomes.

I prefer to use both methods of measurement: absolute number changes matter for the most popular taxons, and relative amounts matter more for the less-popular ones.

## *Comparison over time*

### **All my uBiome results in a single table**

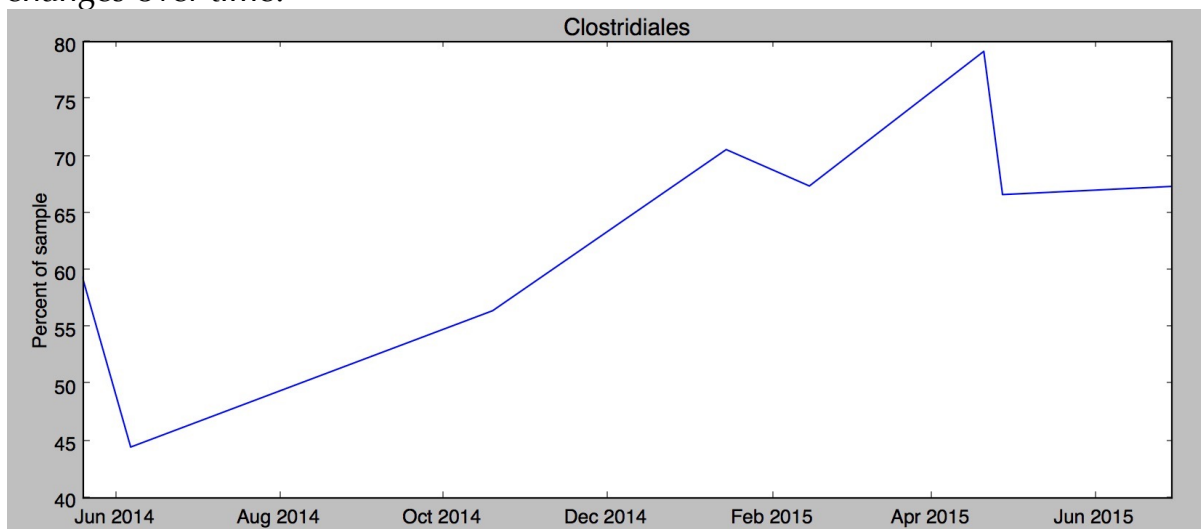
I often read news about a fresh scientific discovery involving the microbiome and immediately wonder if the discovery applies to me. For example, I recently saw [a study from Oregon State University](#) that seemed to find a link between high sugar diets and “cognitive flexibility”, i.e. your ability to adapt and adjust to changing circumstances. The study’s author, Kathy Magnusson, a professor in the OSU College of Veterinary Medicine, found that mice who eat lots of sugar have elevated levels of Clostridiales bacteria, and that this seemed to relate to a slower ability to solve a maze. Hmmm, I thought — how much Clostridiales do I have?

If you have just one uBiome result, that’s easy: log into <http://app.ubiome.com> and search for it in the section “All My Bacteria”. (As far as I know there’s no “search” button yet on the uBiome dashboard). But in my experience a single result doesn’t tell you much. You really need at least two and hopefully several uBiome results to see what might be actionable. In my case, I want to know how my Clostridiales may have changed over time.

I programmed a short Python script to generate a single Excel table with every bacteria I’ve ever found, and then a series of columns with the amount found in each sample. Something like [this](#):

Home Insert Page Layout Formulas Data Review View Table								
A262		Clostridiales						
	A	B	C	D	E	F	G	H
1	tax_name	tax_rank	5/1/14	6/1/14	10/1/15	1/1/15	2/1/15	4/1/15
261	Clostridiaceae bacterium DJF_VR07	species	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
262	Clostridiales	order	59.42%	44.51%	56.49%	70.62%	67.42%	79.24%
263	Clostridiales bacterium 60-7e	species	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%
264	Clostridiales bacterium A2-162	species	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
265	Clostridiales bacterium CIEAF 022	species	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
266	Clostridiales bacterium L2-14	species	0.23%	0.00%	0.03%	0.57%	0.01%	0.08%
267	Clostridiales bacterium oral clone MCE3_9	species	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%
268	Clostridiales bacterium RM3	species	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
269	Clostridiales Family XI. Incertae Sedis	family	0.00%	0.00%	0.03%	0.04%	0.00%	0.00%

The data makes it easy to generate a chart showing how my Clostridiales changes over time:



Hmmm, in my case it looks like something happened since last fall to increase my Clostridiales levels. Maybe it was the potato starch I tried in order to [hack my sleep](#)? Was it my [trip to Central America](#) in February? And of course the biggest question: has the increase affected my cognitive flexibility? I'm [not really sure](#). Whatever happened, the level of Clostridiales seems to have stabilized in the past couple of months.

uBiome has identified more than 900 unique taxa (groups of organisms) in the half-dozen samples I've submitted over the past year, and after running this script I have them all laid out on a single page. Now, armed with this one spreadsheet I can search anytime for a new microbe and quickly see if I have it now, or if it's ever been detected in a previous test. Reading news about microbiome has taken on a whole new personal meaning when I can see if the discovery relates to me.

If you know a little Python, you can make the same spreadsheet with your samples using the ubiome.py Python module on the [ubiome-opensource GitHub repository](#); the script that generated my spreadsheet is [there](#) too as an example. And while you're at it, please upload your own uBiome sample results to the same repository so we can compare.

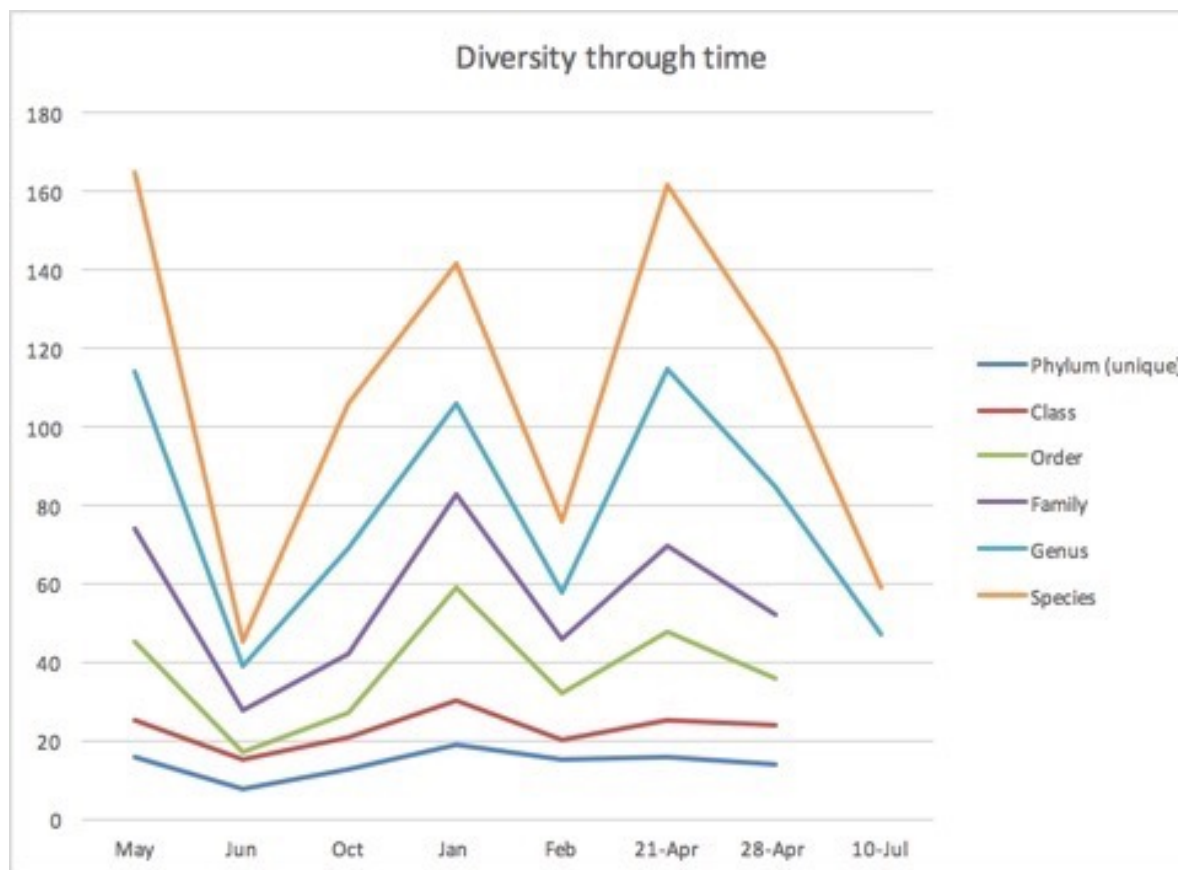
## Measuring Diversity

### Diversity

While there are few agreed-upon standards for what constitutes a good or bad microbiome, nearly everyone agrees that a *diverse* microbiome is better than one that is less diverse. How can we measure diversity?

Ecologists have been interested in this question for a long time and they've developed several metrics to describe how diverse a particular ecosystem is compared to another. This section will show how to apply these ecology metrics to your microbiome.

You could simply count all the different species (or genera or phyla) in a sample and track that over time. The more unique species, the more diversity. Here's what this chart looks like for me:



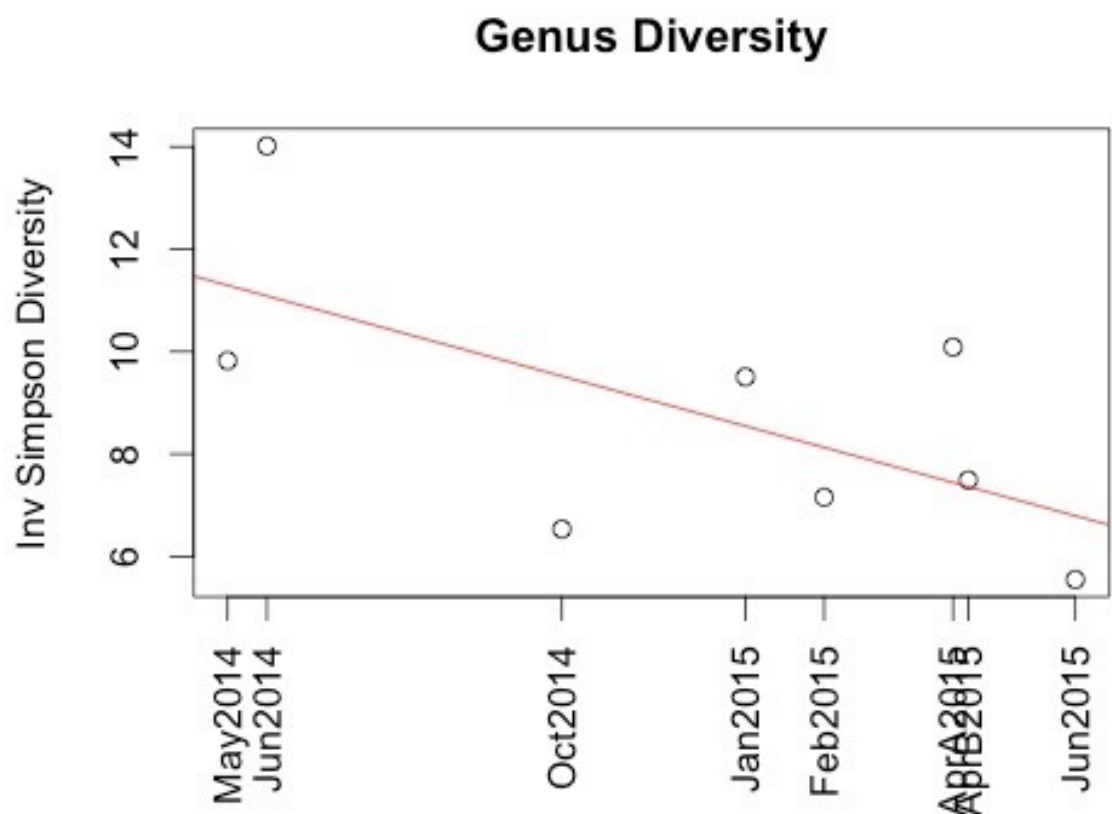
Since 16S technology doesn't capture all the species information, or the genera or phyla information for that matter, a simple count of the number of organisms detected is not terribly useful. In my case, uBiome identified between 90-97% of all the phyla in my samples, but only between 49-51% of the species. That makes apples-to-apples comparisons difficult to interpret.

Ecologists have suffered from this problem for a long time, and they came up with a few metrics to get around the problem. They start by considering what it means to say something is more diverse than another. Consider a forest that has 1,000 trees in it. If all the trees are, say, aspen trees, then that forest is not as diverse as another one that also has 1,000 trees and 1,000 unique species.

Ecologists call this the Shannon number, after the information theorist Claude

Shannon, who was the first mathematician to systematically try to measure information. To Shannon, whose work was concerned with code breaking in World War II, a radio signal that carries information (i.e. a code) is noticeably different from one that is random noise. He applied a specific formula to tell how different a signal looks compared to random noise, a variation of which can be applied to an ecosystem to tell how different it is from one that is completely dead (0) or has nothing but the same or similar organisms.

There is a slightly more ecologically interesting version of the Shannon number, called the Inverse Simpson number, that looks at the total number of unique life forms in an ecosystem and then weights each by the number of individuals of that type of species. Here is my diversity as measured with the Inverse Shannon number:



As you can see, the overall level of diversity in my samples has been going down over the past year. Why?

[Reasons why this might be, next steps, etc]

## **CHAPTER FOUR**

### ***Experiment Examples***

## ***A/B Testing***

What is A/B testing?

[Here's a discussion of the science behind A/B testing, why it's important to compare apples to apples, etc.]

### **Example Experiments**

The best way to start is to experiment on yourself. Here are a few examples:

[Note: in this draft these examples are mostly unedited from blog posts. Need to update them and make them fit the rest of the book, but are here as examples to give the basic idea]

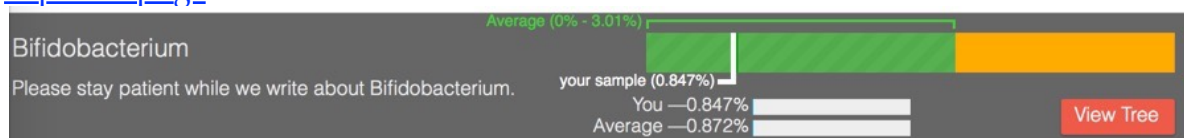
## Hacking my sleep

### Hacking my sleep

About a year ago, I started to notice my sleep becoming less regular. Nothing serious — thankfully, I've never had sleep problems — but I was waking up too early, and I didn't seem to be quite as refreshed. Maybe it's just a sign of age, I thought, until I read in Martin Blaser's excellent book *Missing Microbes* (p.304) that most (80%) of the precursors to the sleep- and mood-regulating neurotransmitter serotonin [is made in the gut](#). Could my gut microbes affect my sleep?

A few internet searches later led me to evidence that [B. infantis modulates tryptophan](#), the stuff in turkey that urban legends have long blamed for [that sleepy feeling](#) you get after Thanksgiving dinner, so my first step was to look at my gut biome results to see my levels of bifidobacterium.

As I suspected, I had no B. infantis, and in fact my overall bifido numbers were pretty low as you can see from this item on my uBiome results [sample explorer page](#):



You may already know about pre-biotics, foods that feed bacteria, as opposed to pro-biotics, which are simply pills or foods that already contain a bunch of (presumably) beneficial microbes. Lately a number of people have noticed that a particular kind of starch, so-called [resistant starch](#) is a prebiotic that acts like a yummy smorgasboard for bifido and other bacteria. I followed [a protocol](#) that uses plain ole [Bob's Red Mill Potato Starch](#) (easy to find any nice grocery store): just a couple of tablespoons one hour before bedtime to give the starch time to make it to the upper colon where the bifido live.

Whoah! You wouldn't believe how wonderfully I slept that night. Over 8 hours of rock-solid, uninterrupted sleep, and more vivid dreams than I've had in years. It was amazing!

After a few days of this, I submitted another sample to uBiome for testing:



Holy [bloom](#), Batman! That's just plain stratospheric: up from 0.847% before the potato starch, to over 5.87% afterwards. That 8x improvement it seems clearly explains my much-improved sleep.

But nothing's free, right? If the bifido are increasing that much, then something else is decreasing. My sleep might be improving, but am I sacrificing some other aspect of my long-term health?



To find out, I wrote a [couple of utilities](#) to examine my uBiome results in more detail. One of the things I like most about uBiome is that they give users access to the raw data, which [you can convert](#) to work in Excel or another programming environment. If you're interested in more details, you can [read how I did this](#) later, but here are some of the major species that went extinct after I took potato starch:

```
##    missing.count_norm
missing.tax_name
## 1                8295        Bifidobacterium
tsurumiense
## 2                4650
Subdoligranulum variabile
## 3                2074        Dialister sp. oral
clone BS095
## 4                780 Desulfovibrio sp. oral
clone BB161
## 5                475        Adlercreutzia
equolifaciens
## 6                459
Ruminococcus sp. ID1
```

What do these microbes do? After a few more internet searches, my answer is: nobody knows! We have some guesses, but so far scientists just haven't had enough samples from real people to understand much. That's another reason I hope you'll submit your samples to uBiome, to increase the number of data points that scientists can work with in hopes of understanding this better.

Meanwhile, my sleep continues to be much better than before I began this experiment, so I've stopped taking potato starch and have been substituting a few other foods that seem to affect the microbiome.

## *Visit the jungle*

### **My microbiome in the jungle**

How much does travel affect your microbiome? In a famous experiment [published in 2014](#), Duke University scientist Lawrence David tracked the daily microbiomes of two people for an entire year and [found significant differences](#) when one of the people travelled outside the U.S. Would the same thing happen to me?

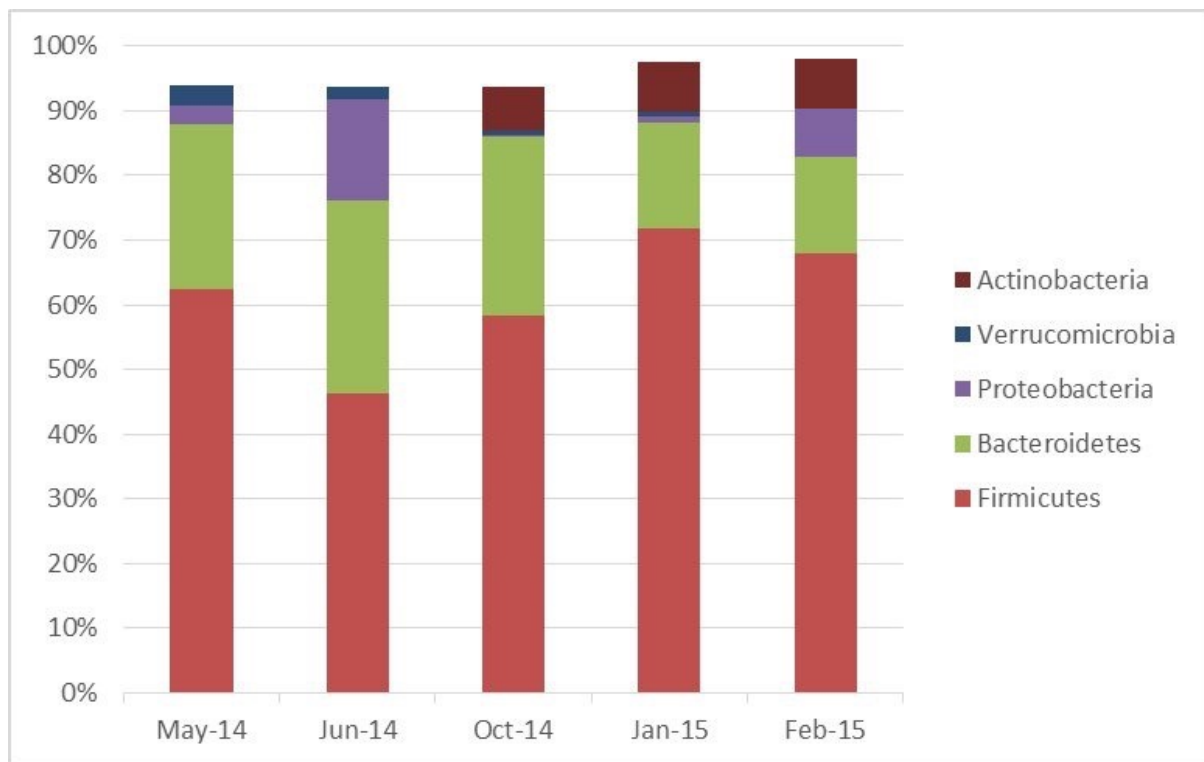
According to my latest uBiome results, the answer is yes. I recently travelled from my home in Seattle to Central America to celebrate my wife's birthday. We spent most of our time in a rural, jungle part of Belize, about a half hour's drive from Benque, near the border with Guatemala. Besides viewing the fantastic, well-preserved Mayan ruins, we also did horseback riding, cave exploration, and of course plenty of eating.

Here's a selfie I took in front of the incredibly well-preserved thousand-year-old pyramid at [Tikal](#):

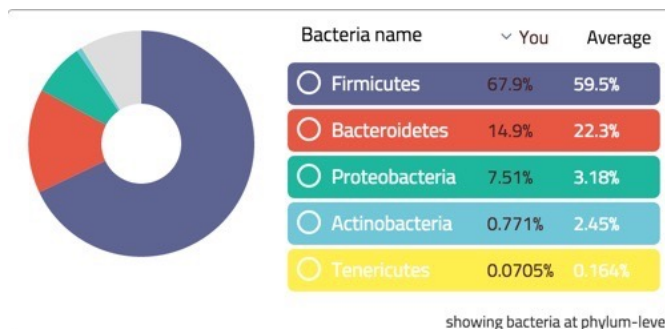


I submitted one uBiome gut kit test just before leaving on our trip, and another immediately after returning home. Since I've been testing myself regularly with uBiome for the past year, I have a good idea of what [my "normal" gut biome](#) looks like. Here's the overall picture through time, including results from my [sleep hacking experiments](#) between October and January, which resulted in those big chunks of Actinobacteria, colored brown in the graph, now

mostly gone:



Now let's look more closely at the latest sample, the one taken right after returning from the jungle:



The first thing to note is the big increase in the ratio of firmicutes to bacteroidetes – a change which is often [associated with obesity](#). Interestingly, I did gain a couple of pounds during the trip, maybe from all that tasty coconut rice.

Otherwise my uBiome results point in a few more interesting directions:

- Diversity: Oddly, my gut biome diversity went down slightly. Before the trip, uBiome found 19 unique phyla. Afterwards, there were only 15. You wouldn't normally expect diversity to drop after exposure to novel microbes from the jungle. But I think there's an easy explanation:

- Increase in unidentified organisms: uBiome was able to identify only about 91% of what it found at the phyla level. In my previous tests, they found closer to 95+%. Maybe my apparent drop in diversity was simply a drop in identifiable bacteria.

- Pathogen plunge: Counts of my “bad” bacteria, including members of the notorious *Clostridium* genus, which includes many nasty species (e.g. the infamous antibiotic-resistant *C. Difficile*) disappeared, dropping from 0.66% to 0.18%.

- Bifido back to normal. The bifidobacterium that increased 10X when I started taking potato starch hoping to improve sleep, went back down to normal levels. My sleep quality, by the way, hasn’t changed.

- Akkermansia is gone. Usually considered a “good” bacterium, my levels had been dropping since last Summer, and now finally disappeared completely. Hopefully this is just temporary.

Using my [free public uBiome analysis tools](#), I see that I picked up the following genera while in the jungle:

```
head(newAfterBelizeGenus)
##   missing.count_norm missing.tax_name
## 1                13750   Pectobacterium
## 2                 1190     Serratia
## 3                 970    Enterobacter
## 4                 749    Planifilum
## 5                 529   Enterococcus
## 6                 441   Escherichia
```

and a few that went extinct:

```
head(extinctAfterBelize)
##   missing.count_norm missing.tax_name
## 1                 6269   Akkermansia
## 2                 5619   Corynebacterium
## 3                 1702   Butyrivibrio
## 4                 1257   Arthrobacter
## 5                 881   Turicibacter
## 6                 787 Hydrogenoanaerobacterium
```

What do these bacteria do? As always, who knows? The science is so new that when I look up most of these genera online and elsewhere, I find almost nothing relevant. All the more reason I hope you’ll submit a sample to uBiome the next time you travel in the jungle, so we can compare and learn together.

## Looking into your mouth

### Looking into my mouth microbiome

The gut biome is interesting enough, but bacteria colonize just about every part of the body, so recently I've been studying my uBiome mouth test results. The simple [GitHub RuBiome utilities](#) I use for analyzing my gut will work for that too, so here's a short example of how I did it:

First I [loaded my uBiome data](#) into two variables, one for each sample: June 2014 (junMouth) and the other from October 2014 (OctMouth), after [a visit to my dentist](#).

Let's see which species are new in the later (October) sample:

```
octToJunChange <- uBiome_sample_unique(OctMouth,junMouth)
##   count                               missing.tax_name
## 1  3640                               bacterium NLAE-z1-P562
## 2  2725                               Streptococcus sanguinis
## 3  2075                               Capnocytophaga gingivalis
## 4  1969 Peptostreptococcus sp. oral clone FG014
## 5  1618                               Granulicatella adiacens
```

One of those species, *Streptococcus sanguinis* looks interesting. [Wikipedia](#) says this:

*S. sanguinis* is a normal inhabitant of the healthy human mouth where it is particularly found in dental plaque, where it modifies the environment to make it less hospitable for other strains of *Streptococcus* that cause cavities, such as *Streptococcus mutans*.

No cavities? Nice! More good news: this quick check confirms that I don't have any *S. mutans*:

```
OctMouth[grepl("Streptococcus",OctMouth$tax_name),]$tax_name
## [1] Streptococcus                               Streptococcus
##    pseudopneumoniae
## [3] Streptococcus sanguinis                       Streptococcus constellatus
## [5] Streptococcus anginosus group                 Streptococcus sp. oral
##    clone GM006
## [7] Streptococcus thermophilus                   Streptococcus oralis
## [9] Streptococcus gordonii
## 250 Levels: [Eubacterium] sulci ... Veillonellaceae
```

Then I look at the species that disappeared (went extinct) between the two samples:

```
junToOctChange <- uBiome_sample_unique(junMouth,OctMouth)
##   count                               missing.tax_name
## 1  6034                               Capnocytophaga granulosa
## 2  4153 Peptostreptococcus sp. oral clone FL008
## 3  3375                               Prevotella sp. oral clone ID019
## 4  2691                               Streptococcus rubneri
```



## 5 1571

Prevotella buccae

Anything in the genus [Capnocytophaga](#) is an opportunistic pathogen, so I say good riddance. Usually they're fine, but if your immune system dips they can turn bad.

[Streptococcus rubneri](#) was discovered a couple years ago, so little is known about it.

[Prevotella buccae](#) is more interesting. It seems to be implicated in periodontal disease (yuk!) but that genus is involved too in breaking down proteins and carbohydrates.

Hard to say what's really going on. Meanwhile, here are the biggest changes (increase) since the first sample:

```
junToOctCompare <- uBiome_compare_samples(junMouth,OctMouth)
```

##		tax_name	count_change
## 64	Streptococcus	pseudopneumoniae	62007
## 68	Veillonella	sp. oral taxon 780	8065
## 41	Neisseria	oralis	4693
## 2	Abiotrophia	sp. oral clone P4PA_155 P1	2308
## 28	Granulicatella	elegans	1987

Whoah! That first one, Streptococcus pseudopneumoniae, looks nasty! [Wikipedia](#) says it may cause pneumonia, though a recent medical journal says more hopefully that it [“treads the fine line between commensal and pathogen”](#) which is a scientific gobbledygook way of saying nobody has a clue. All the more reason to keep testing, submitting, and getting more data.

## Comparing with food

### How does food affect my microbiome?

I recently tested two samples each taken a week apart. For several weeks before and during this period, I carefully tracked all the food I ate using a free app from <http://myfitnesspal.com>, hoping to get some insights into how diet affects my own microbiome. Here's how I used these utilities to conduct the analysis.

Begin by setting my working directory using the R command "setwd" or from within the R environment. Then load the utility files using the R "source" command.

```
setwd("uBiome/sources")
source("uBiomeCompare.R")

# set your working directory to the place where you keep your data.
setwd("/uBiome Results")

A<-read.csv("sprague-ubiome-150421.csv")
B<-read.csv("sprague-ubiome-150428.csv")
```

How do the samples compare? I'll create a new variable A\_vs\_B and assign it to the results of running the uBiome\_compare\_samples function.

```
A_vs_B<-uBiome_compare_samples(A,B, rank="species")
```

```
head(A_vs_B)
##                tax_name count_change
## 1      [Ruminococcus] obeum      -6949
## 2  Acidaminococcus fermentans        112
## 3 Adlercreutzia equolifaciens      -637
## 4      Alistipes finegoldii       1421
## 5      Alistipes indistinctus         19
## 6      Alistipes onderdonkii       1017
```

By default, uBiome\_compare\_samples returns a result sorted alphabetically. Here's how to sort by species count. For brevity, I'll show only the top and bottom ten values:

```
sortAB <- A_vs_B[order(A_vs_B$count),]
head(sortAB,10)
##                tax_name count_change
## 54      Coprococcus sp. DJF_CR49      -25077
## 81 Peptostreptococcaceae bacterium TM5      -23030
## 71      Methanobrevibacter smithii      -22461
## 43      Clostridium clostridioforme      -19006
## 8      Anaerostipes hadrus      -18478
## 88      Ruminococcus bromii      -17735
## 31      Blautia faecis      -13980
## 41      Clostridium baratii      -13565
## 1      [Ruminococcus] obeum      -6949
```

```
## 29          Bifidobacterium animalis          -6455
tail(sortAB,10)
##          tax_name count_change
## 33      Brachyspira sp. NSH-25      2952
## 27      Barnesiella intestinihominis      5577
## 79      Parasutterella excrementihominis      5641
## 30      Bilophila wadsworthia      5819
## 85      Roseburia inulinivorans      5898
## 69      Lactobacillus rogosae      9910
## 26      Bacteroides uniformis      12303
## 86      Roseburia sp. 11SE38      25807
## 61      Faecalibacterium prausnitzii      59182
## 14      bacterium NLAE-zl-P430      66457
```

Negative numbers indicate a drop from the first sample; positive numbers mean there are more in the second sample than in the first. The numbers on the left are references to the original R data frame row before the sort.

As noted above, all references to count or count\_change in these examples are using uBiome's normalized count, which you can think of as units of parts per million. Dividing by ten thousand gives the percentage.

In this example, I see a large increase in *Faecalibacterium prausnitzii*, often considered a marker for health. I'll want to look at my meals over the test period to see what might have driven this positive change.

Using the macronutrient data I collected daily with MyFitnessPal (exported to CSV with the handy exporter at <https://www.designbyvh.com/myfitnesspal-export-data/>) I produced this simple chart to see if my eating habits had an effect on my microbiome:

Date	Calories	Carbs	Fat	Protein	Cholesterol	Sodium	Sugars	Fibre
Average (month)	1841.7	192.2	102.7	94.9	268.0	2298.5	64.1	15.2
Average (test period)	2242.6	241.9	124.7	108.7	262.3	2814.3	78.3	16.9
Difference from Ave	400.9	49.7	22.0	13.9	-5.7	515.7	14.2	1.6
% Diff from Ave	122%	126%	121%	115%	98%	122%	122%	111%

As you can see, though I apparently ate about 20% more overall during the week of my test, my dietary cholesterol was conspicuously lower than normal. Did this affect my microbiome? My next step will be to search the medical and biology literature to see if there are known relationships between dietary cholesterol and *Faecalibacterium*. Will I find something new? That's the exciting part about this type of citizen science: armed with my own data and imagination, new discoveries are everywhere.

## ***Beyond the Microbiome: DNA***

### **How to analyze your 23andme results**

[This section describes briefly how to pull your 23andme results to your computer for analysis and comparison with microbiome results]



23andme says I'm a slow caffeine metabolizer because I have the AC genotype at SNP rs762551. You'd think that means I'm extra sensitive to caffeine before bedtime, but that's not the case: I sleep just fine even if I have a cup of high-octane coffee after dinner. My 99-year-old grandmother drinks coffee by the potful, crediting its warmth as a calming effect to make her drowsy.

A study in *Nature* by Javier A. Ceja-Navarro et al names *Pseudomonas Fulva* as a bacterium active in the guts of a coffee bean pest. Caffeine is normally toxic to insects, but *P. Fulva* neutralizes the caffeine, apparently using the demethylase *ndmA* gene.

Maybe my own caffeine metabolism is also affecting by a similar gut bacteria that I have in abundance but which is missing in other people. I already checked for *Pseudomonas Fulva* — I don't have it in any of my uBiome samples. I do have abundant levels of the genus *Pseudomonas*, but that is not the same thing.

It should be possible to screen every one of my gut bacteria demethylizing *ndmA* gene, but I'm not sure how to do that. If I did find the gene in one of the gut bacteria I harbor, then that would be pretty cool: a microbe that helps me drink coffee.

The following code, written in R, will let you pull all your 23andme results into a data frame that you can analyze in more detail.

```
> g e n e s < -  
read.delim("genome_Richard_Sprague_20080928082410.txt",comment.char="#"  
)  
> head(genes)
```

will list all the genes, now presented in a dataframe with headers:

```
> names(genes)  
[1] "rsid"      "chromosome" "position"  "genotype"  
>
```

Now find a particular SNP using something like this:

```
> genes[genes["rsid"]=="i3001542",]  
      rsid chromosome position genotype  
579746 i3001542      MT   16522      CC
```

## **CHAPTER FIVE**

### ***Organisms to Watch***

## ***Good and Bad***

As tempting as it is to categorize bacteria into groups like “good” and “bad”, you should avoid that mistake. Very few organisms are always and everywhere either beneficial or harmful. Everything depends on the circumstances.

For example, from [Blaser p119]

"viridans" streptococci "live peacefully in everyone's mouth...a leading cause of heart valve infections...normal residents of the mouth, only occasionally entering the bloodstream and landing on a previously damaged valve....if we mix harmless Viridans with pathogenic Group A strep, viridans always win out. They knock back the strep."

Other examples:

H.Pylori

## ***Gut Guardians***

### **The Gut Guardians**

Grace Liu is a PharmD from the Bay Area who has been fascinated by the gut microbiome for many years. She focuses on the following organisms she calls the “gut guardians”, the keystone species that, when properly groomed will keep pathogenic organisms at bay and ensure the overall health of the body’s ecosystem.

### **The importance of butyrate**

The cells lining your intestines will die without proper nourishment, and a key food they expect is butyrate, a type of short-chain fatty acid. Fortunately, your colon is covered with bacteria that can ferment plant fiber, producing butyrate as a by-product.

From Wikipedia:

Butyrates are important as food for cells lining the mammalian colon (colonocytes). Without butyrates for energy, colon cells undergo autophagy (self digestion) and die.[1] Short-chain fatty acids, which include butyrate, are produced by beneficial colonic bacteria (probiotics) that feed on, or ferment prebiotics, which are plant products that contain adequate amounts of dietary fiber. These short-chain fatty acids benefit the colonocyte by increasing energy production, and cell proliferation and may protect against colon cancer. [2]

Butyrate is a major metabolite in colonic lumen arising from bacterial fermentation of dietary fiber and has been shown to be a critical mediator of the colonic inflammatory response.

### **Faecalibacterium prausnitzii**

<http://www.sciencedirect.com/science/article/pii/S1369527413000775>

*Faecalibacterium prausnitzii* is the most abundant bacterium in the human intestinal microbiota of healthy adults, representing more than 5% of the total bacterial population. Over the past five years, an increasing number of studies have clearly described the importance of this highly metabolically active commensal bacterium as a

component of the healthy human microbiota. Changes in the abundance of *F. prausnitzii* have been linked to dysbiosis in several human disorders.

**Roseburia**

This genus, under the phylum *firmicutes*, is known to be associated with health.

Christensenellaceae

Christensenella

Akkermansia

Bifidobacteria longum

Bifidobacterium

B. Longum as % of total Bifido

Bacteroidetes/Firmicutes ratio

## **CHAPTER SIX**

### ***Next Steps***

## ***Fermented food***

Once you've developed an appreciation for the importance of the microbiome, you'll want to become more of a farmer, growing your own symbionts.

People have been fermenting food since the distant past, making everything from beer to cheese. Fermentation is an ideal way to preserve food beyond the date at which it is practical to eat.

Fermenting your own foods is surprisingly easy.

### **Yogurt**

One of the easiest ferments is home-made yogurt. Pour a few cups of whole milk into a saucepan and heat to not-quite-boiling. You'll need to get it above 160 degrees, the temperature that kills bacteria in milk, but if it goes all the way to boiling the milk will taste scalded. Even that's not the end of the world, but you may as well try to keep it under that. Use a kitchen thermometer the first few times if you like, but soon you'll be able to tell intuitively because you'll see steam rising from the surface of the milk, but no boiling bubbles.

After the milk has been heated, remove and pour it into a glass bowl. Let it sit until it cools to about 110 degrees. Here, a kitchen thermometer is more useful, especially the first time because the exact temperature is more important. Warmer than 110 degrees and the *lactobacillus* yogurt bacteria will die; too much lower and they won't grow. Once you have the hang of it, you'll be able to tell based on the touch: it'll feel unpleasantly warmer than your hand (which should be 98 degrees) but not too warm to touch.

Once the milk hits that rough temperature zone, pour in a tiny amount of yogurt from another source. This is called the "starter" yogurt and the best is to use a bit of leftover from your previous batch, but for your first attempt you can use any commercially-purchased yogurt. Sweetened, flavored, non-fat — any kind is okay as long as it has live cultures in it.

Be sure not to use too much yogurt. Definitely no more than a tablespoon, and perhaps even just a teaspoon or two. Because of the small amounts, you'll be tempted to think you need more, but in this case more is definitely not better. Too much starter will suffocate the mixture. Whenever I've had an unsuccessful batch it's almost always due to having too much starter.

Mix the starter thoroughly into the lukewarm milk and then cover it up and put it someplace warm. The key is to keep the mixture close to that magical 110 degree point. Some people put it overnight in an oven set to the lowest temperature, but if you don't like running the oven overnight you can try wrapping it tightly with something that will retain the heat through the night. If have a high-quality thermos-style cooler, you can set it inside there, and some people cover it with thick towels and set it in the warmest part of the house. Another option is to use a heating pad, set to the lowest setting and placed underneath the bowl. The important thing is to keep the mixture at the warm

temperatures that the bacteria need in order to breed.

Set in a quiet place for about 8-10 hours. Overnight is ideal. When you open the bowl again, you'll find the mixture no longer resembles milk, but has been transformed into a much thicker consistency — a yogurt. It will taste sour, and if you're not accustomed to plain, unsweetened yogurt you may even think that is has spoiled. Nope, that's just how pure, real yogurt is meant to taste.

### **Pickling**

The next step after yogurt-making is also easy. It takes a little longer, from one to several weeks, but lets you expand your fermentation skills to many more foods. The most basic one is cabbage, to make your own sauerkraut.

Think of fermentation as a type of farming, only instead of large plants that you see, you're farming with invisible microbes. In both cases the object is to start with a small seed that grows into a full crop. For sauerkraut, the microbes are naturally occurring on the leaves of plants and vegetables. You just need to put them into the right environment so the microbes can grow.

Start with a small head of cabbage and a grater. If you don't have a grater, then just chop it into small pieces. Cabbage is already covered with tiny microbes, but the interesting ones will not reproduce in the presence of oxygen; you need to get them out of the air. The best way to do that is to submerge them in a liquid — in this case we'll use the naturally occurring water inside the cabbage leaves themselves. The purpose of grating or chopping is simply to let the water escape from the leaves until it covers the entire mixture.

You'll need some type of container to submerge the leaves. Best is something made of glass or ceramic, though any container is fine as long as it doesn't react with the acids that will be produced by the microbes. A standard mason jar works, for example, or even a cleaned-out jar of pasta sauce. Make sure you have a lid for the container, and make sure the lid doesn't close too tightly. If you use something with a screw-top lid (like a pasta sauce jar) just pop a few tiny holes in the lid. The microbes will be producing carbon dioxide as they ferment, and you'll need a way for the gases to leave or the container will explode.

Begin packing the cabbage leaves into the container as tightly as possible. While doing this, most people add some salt. The salt serves two purposes; it adds flavor, and it helps draw more water out of the plant. Just a small amount of salt is enough: about what you would use if you were seasoning it for flavor: a few shakes on each cabbage leaf as you peel it is enough, or maybe a total of a tablespoon or two for a small head of cabbage. Continue to pile more leaves into the container until it is full.

If you're doing everything correctly, there will be a nice layer of liquid that completely covers the leaves in the container. Remember that the microbes can't reproduce if there is air, so covering with liquid is essential. If you see any green leaves poking out of the liquid, pack them down further. It may help to find an insert of some kind to place directly over the leaves, perhaps anthe lid to a smaller sized jar, one that fits inside the container and lets you squish the leaves even tighter. The key is to keep everything submerged in liquid.



Put as many cabbage leaves as you can into the container and close the lid, but be careful: the bacteria in the jar will soon start to produce carbon dioxide gas, which will put pressure on the container. That's the reason for the air holes you poked in the lid beforehand. You may want to cover the entire jar with a loose-fitting piece of cellophane, or an old cloth, as a precaution in case the aroma begins to attract insects.

That's it! Leave the fermenting jar in a cool, undisturbed place and then check on it for the first few days. Make sure the liquid has entirely submerged the cabbage leaves. Within a day or two you'll also notice some air bubbles in the liquid — the sign that the fermentation has begun to work. Now, just let the jar rest for another several days or more. The longer it sits, the more acidic it will become. If you like a sour taste, let it sit for a week or two, maybe even up to a month or more. If you want a more subtle flavor, take it out in a few days. Feel free to sniff the jar every day or two to decide when you think it's ready.

When finished, it will look and smell like sauerkraut. If some of the leaves were not entirely submerged in liquid, it's possible that a layer of moldy fungus has appeared, unsightly but not a problem. Just scrape off whatever doesn't look like sauerkraut and eat the rest.

### **What's happening in sauerkraut**

The transformation from cabbage leaves to sauerkraut happens thanks to a complex interaction of bacteria starting with some members of the *Leuconostoc* species that occur naturally on the cabbage leaves. These are anaerobic organisms — they can't live in oxygen. When submerged in the cabbage liquid, however, they thrive, consuming all the sugars they can find and converting them to lactic acid, giving off carbon dioxide as a byproduct. You'll know the reaction is occurring when you spot tiny bubbles at the surface of the liquid a few hours after beginning the reaction.

Sadly for the *Leuconostoc*, they are too good at what they do. Eventually they'll consume all of the sugars they find transforming the liquid into a sour concoction, and as result the environment turns too acidic for them and they die.

By then, the conditions have become ideal for another species, *lactobacillus*, which was pretty much designed for acidic liquids, and they finish off the remaining sugars, until the acidic level reaches PH=3, at which point their work is done.

Fortunately for humans, the pathogens that could cause us harm are unable to survive in this high-acidic environment, and would be out-competed by the existing microbes anyway, so the final mixture is perfectly healthy.

### **More fermentation**

Just about any food can be fermented, as people in all cultures have known since pre-history. Many common ethnic foods, such as miso (soybean) or kimchee (cabbage), are possible only thanks to fermentation. When you eat "cured" bacon or sausage, you are eating something that resulted from a complex interaction between microbes and meat.



### ***Overselling the microbiome***

Now a warning. Our understanding of the microbiome is in its infancy, with new techniques and discoveries coming along every day. The costs have plummeted dramatically, whether measured in money or in the time and resource commitment needed to explore the science. After all, this book is itself proof for how easy it is for “normal” people can learn and make their own discoveries.

All of this new interest is a recipe for getting carried away, and you’ll find many examples, especially once you understand the science a bit for yourself. Jonathan Eisen is a microbiologist who has been studying the microbiome for decades and he has created a blog list he calls “Overselling the Microbiome Award”, with many examples.

## ***Annotated References***

Here are my suggestions for additional reading.

Read more about Toxoplasma Gondii: <http://www.theatlantic.com/magazine/archive/2012/03/how-your-cat-is-making-you-crazy/308873/>

This is a new field, so much of the written literature remains in academia

### **Books**

*Missing Microbes* by Martin Blaser

*The Wild Life of Our Bodies* by Rob Dunn

*Plague Time: The New Germ Theory of Disease* by Paul Ewald

*An Epidemic of Absence: A New Way of Understanding Allergies and Autoimmune Diseases* by Moises Velasquez-Manoff

Practical

*Wild Fermentation: The Flavor, Nutrition, and Craft of Life Culture Foods*  
Sandor Kaz

### **Authors**

Carl Zimmer

### **Magazine Articles**

GERMS ARE US

Bacteria make us sick. Do they also keep us alive? BY MICHAEL SPECTER

[http://www.newyorker.com/reporting/2012/10/22/121022fa\\_fact\\_specter?](http://www.newyorker.com/reporting/2012/10/22/121022fa_fact_specter?currentPage=all)  
[currentPage=all](http://www.newyorker.com/reporting/2012/10/22/121022fa_fact_specter?currentPage=all)

The Atlantic: How Your Cat Is Making You Crazy (Feb 6, 2012) <http://www.theatlantic.com/magazine/archive/2012/03/how-your-cat-is-making-you-crazy/308873/>

### **Web Sites**

<http://mrheisenbug.wordpress.com/>

<http://www.bacteriamuseum.org/>

### **Courseware**

Gut Check: Exploring Your Microbiome

## Other

Foundation for Bacteriology (which, according to the Guidestar list, has not filed reports for years and may no longer exist)

### Diet and nutrition

*The Paleo Manifesto: Ancient Wisdom for Lifelong Health* Durant, John

The best summary so far of the motivation and principles of ancestral health. The author is a student of Steven Pinker's, from Harvard, and writes with a general, more academic orientation rather than as a how-to manual. The basic principle, that the modern world is not our natural habitat, makes much sense, and I like the way he applies that rule to diet and exercise, plus sleep and much more.

*Paleofantasy: What Evolution Really Tells Us about Sex, Diet, and How We Live* Zuk, Marlene

Although it tries to debunk ancestral diets with proof that ancient humans had too much variation, this book targets mostly a straw man caricature of the paleo movement, and rarely addresses the real issues.

*Why We Get Fat: And What to Do About It* Taubes, Gary

A well-researched, easy to read, but thorough discussion of obesity that concludes that carbohydrates, not calories, are key. The simple, seemingly obvious belief that a person's weight is a function of "calories in and calories out" will seem much less obvious and mostly wrong by the end of this book.

*All Natural: A Skeptic's Quest for Health and Happiness in an Age of Ecological Anxiety* Johnson, Nathanael

Although frustratingly equivocal with recommendations, I liked the survey of the advantages and disadvantages of "mainstream" versus "alternative" approaches to health, on everything from childbirth, vaccinations, and raw milk.

*Eating on the Wild Side: The Missing Link to Optimum Health* Robinson, Jo

A highly practical summary of fruits and vegetables: which are good for you and why. Every page is full of interesting, often counter-intuitive tips to eat more healthily. Examples: frozen blueberries are just as healthy as fresh, but broccoli loses most of its nutrition within hours after picking. Carrots cooked with butter are much healthier than raw.

*In Defense of Food: An Eater's Manifesto* Pollan, Michael

If you're confused about diet, this is the best advice yet. "Eat food. Not too much. Mostly plants".

### Far out theories

*Plague Time: The New Germ Theory of Disease* Ewald, Paul

Published in 2002, this book raises the intriguing possibility that most (perhaps all) serious diseases are caused by infections. Certain types of cancers (e.g. HPV) are already known to have viral origins, but imagine how our thinking would change if — when — someday science discovers infectious agents behind other cancers and heart disease. Reading this with other books about the role of microbes has made me far more sensitive to the possibility that science and medicine could one day undergo a huge shift in the way that diseases are diagnosed and treated.

*An Epidemic of Absence: A New Way of Understanding Allergies and Autoimmune Diseases* Velasquez-Manoff, Moises

Another book that explains a provocative idea that our immune systems need regular stimulation by parasites and other infectious agents, or we risk unpleasant side effects in the form of allergies, diabetes, and many other nasty conditions. The remarkable correlation between the hygiene of modernity and the rise of autoimmune diseases makes for powerful evidence that science is far behind in understanding all the consequences of our current lifestyles.

***Endnotes***